CALIFORNIA PRIVACY PROTECTION AGENCY

TRANSCRIPTION OF RECORDED PUBLIC MEETING

MARCH 30, 2022

SACRAMENTO, CALIFORNIA

Present: JENNIFER M. URBAN, Chair
LYDIA DE LA TORRE, Board Member
VINHCENT LE, Board Member
ANGELA SIERRA, Board Member
J. CHRISTOPHER THOMPSON, Board Member
JUSTIN GOURLEY, Moderator
SAFIYA NOBLE, Presenter
GWENDAL LEGRAND, Presenter
CHRIS HOOFNAGLE, Presenter
ANDREW SELBST, Presenter
MARGOT KAMINSKI, Presenter

Transcribed by: Mariam Ayad,
eScribers, LLC
Phoenix, Arizona

--o0o--

TRANSCRIBED RECORDED PUBLIC MEETING

March 30, 2022

MR. GOURLEY: Okay, Chairperson Urban. I think we're okay to start now.

MS. URBAN: Thank you very much, Mr. Gourley. And good morning, everyone joining us today. My name is Jennifer Urban, and I am the chairperson of the California Privacy Protection Agency Board. Welcome back, or welcome for the first time if you didn't join us yesterday, to our March 2022 pre-rulemaking informational sessions.

We are now back in session and this is day two on the program. As a reminder, these sessions are being recorded.

I would now like to ask our moderator, Mr. Justin Gourley, to please conduct the roll call.

MR. GOURLEY: Thank you, Chairperson Urban. I will conduct the roll call now.

Ms. De la Torre.

MS. DE LA TORRE: Present.

MR. GOURLEY: Mr. Le.

MR. LE: Present.

MR. GOURLEY: Ms. Sierra.

MS. SIERRA: Present.

MR. GOURLEY: Mr. Thompson.

Chairperson Urban.

MS. URBAN: Present.

MR. GOURLEY: Chairperson Urban, there are four board members present.

MS. URBAN: Thank you very much, Mr. Gourley. The board has established a quorum. Thank you, board members, appreciate your service in joining us.

Before we get to the substance of the day, as I did yesterday, I am going to go over some of the logistical announcements that are necessary for everyone to be able to follow along as easily as possible. For those of you whom this is a repeat, thank you for your patience as we make sure everyone has a clear understanding.

I'd like to ask first that everyone remember to please check that your microphone is muted when you are not speaking. And please note also that the meeting is being recorded.

Meetings and events involving a majority of board members, including informational and instructional sessions like these, will be run according to the Bagley-Keene Open Meeting Act as required by law.

First, let me sketch the format of the pre-rulemaking informational session so everyone has a sense of how things will proceed today. Each day, yesterday and today, includes a set of experts presentations that

will provide background information to the board, agency staff, and the public on topics that are potentially relevant to the agency's upcoming rulemaking.

If you review the agenda, you'll see that we have an opening item today that I'm going through now and then an item comprising a series of presentations on today's topic. Accordingly, after we finish our item to open the day, we'll do public comment. And then we'll go to the item with our series of presentations for the today.

Let me provide some information on how to engage in public comment. I will call for public comment after each item, so after this introductory item and then after the presentations at the end of the day. Each speaker will be limited to three minutes.

If you wish to speak on an item, please use the "Raise your hand" function, which can be found in the reaction feature on the bottom of your Zoom screen if you want to take a second to locate it. Our moderator will request that you unmute yourself for comment. When your comment is completed, the moderator will mute you.

It is helpful if you identify yourself, but this is entirely voluntary, and you can input a pseudonym when you log into the video conference.

I'd like to remind everyone of the rules of the road under Bagley-Keene. Bagley-Keene does require that

comments be tied to the agenda item in question. Accordingly, please do plan to comment on today's presentations at the end of today's session. Although you will be able to comment on any of the presentations from today, yesterday it was appropriate to comment on the presentations for yesterday. I'd like to remind everyone who speaks in public comment to please stay on topic and keep your comments to three minutes.

Now, a little bit more on the schedule. Yesterday we took a break for lunch and another short break. Today is likely to be similar, excuse me, in terms of breaks, but because of the need to accommodate our guest speakers' schedules, our longer break is most likely going to be a little bit earlier in the day than is traditional for lunch. I just don’t want to surprise people too much.

We expect to break after the first two informational presentations, if you want to check the agenda. We'll also take additional shorter breaks as needed. Please do note that my estimates of timing may not hold. At the same time, I think that staff says to expect to finish today sometime in sort of early to mid-afternoon. It is possible, you know, that it will go longer but that is the expectation.

As I mentioned, this is being recorded. In case you

need to come and go outside of breaks, you won't miss anything that you can't review later.

My thanks to all of the expert speakers who are taking time to present to us today and tomorrow and to all the people working to make the meeting possible. A great deal of work goes into any public meeting or event and this is certainly no exception.

I would like to thank the team from the Office of the Attorney General supporting us today, Mr. Mulai Dajou (ph.) Dajou, who is our meeting counsel, and Mr. Justin Gourley, who is acting as moderator. Ms. Trini Hurtado (ph.) is a conference services expert who organized the meeting infrastructure. And Ms. Stacey Hineson (ph.) is the person who's organizing the administrative staffing and resources.

As I said yesterday, I'd also just like to thank everyone at multiple agencies who have been supporting us, including the Department of Consumer Affairs, the Business and Consumer Services Housing Agency, the Department of General Services, and the Office of the Attorney General, among others.

Before we move to today's presentations, I'd also like to situate today's program with our pre-rulemaking activities and to invite your participation in our pre-rulemaking work. Some of you participated in and some of

you may recall that we started our pre-rulemaking work in the fall of 2021 with an invitation for written comment. And we were delighted to receive many substantive responses to that.

The board also has discussed a desire for informational sessions and informational hearings and that is what we're doing now. We've announced two sets of pre-rulemaking events. First, these informational sessions, and second, stakeholder sessions.

The pre-rulemaking informational sessions today and tomorrow, as I mentioned, will provide background information. The speakers for these informational sessions are academics who study relevant topics and officials from the California Office of the Attorney General, the California Privacy Protection Agency, and the European Data Protection Board. We very much hope that these will be helpful information.

It is important to note that our guest presenters' view should be taken as the view of the agency or the board. They are the presenters' views only.

Our second set of pre-rulemaking events will be pre-rulemaking stakeholder sessions, which will follow in a month or so. Stakeholder session are designed to gather stakeholder input complementary to the written stakeholder input received in response to our preliminary

invitation for comment. Like the written input, this information will be very helpful. There are many knowledgeable stakeholders who can offer input based on their specific experience, their expertise, and so forth.

I got a very helpful question yesterday during public comment, asking me what is a stakeholder. A stakeholder is anyone with an interest in the topics under the jurisdiction of the agency, so Californians privacy. And a stakeholder is anyone from a consumer to a local muni -- a local interest group to a business who's implementing the law to business associations or nonprofits who work on consumer issues, and on and on. It is anyone who is interested in our work.

I also want to be clear about what I mean by experience and expertise. We are hearing today and tomorrow from speakers who have made formal studies of the topics on which they're speaking. I just want to be sure that that doesn't seem to anyone as though that's what we expect for expertise. There are many kinds of expertise.

Stakeholders have many experiences and expertise that will be extremely helpful. For example, as I mentioned, an individual business implementing the California Privacy Protection Act, an individual consumer who has been working to exercise her rights, associations

who represent different groups who have an interest in the rulemaking, and probably many I haven't, you know, added to the list. All of those experiences and expertise will be very helpful, and I really would like to encourage people to consider signing up.

So everyone who is interested in participating, please do consider signing up for the stakeholder session. You can find more information on our website, CPPA.ca.gov, on the regulations page. There, you'll find information about logistics and a link to a sign-up form.

Please note that the date for the stakeholder sessions is not yet set because staff needs to see, you know, how many sign-ups roughly and also to look for -- they're working on venue options that will have a component for people to be able to come in person.

But please even though there isn't a date yet, please do feel free to sign up now. The agency will contact you with options for participation. And you're always free to decline to participate if the final dates are inconvenient for you.

Also, if we get to the stakeholder sessions and you find you didn't have time or didn’t remember to sign up, there will be opportunities for general public comment at those session as well. So please check it out and consider participating. If you have questions, write to

info@CPPA.ca.gov.

All right. Thanks, everybody. And is there any public comment this morning from those in the audience?

MR. GOURLEY: Thank you, Chairperson Urban. It looks like there's a couple. And as a reminder for anybody, please use the "Raised hand" button to indicate that you'd like to make a comment. You will have three minutes to make your comment.

Ms. Gellis -- I hope I said that right -- you now have permission to unmute yourself.

MS. GELLIS: Thank you very much. I'm unmuted I believe. Okay.

Thank you. My name is Cathy Gellis. I'm a lawyer in the San Francisco Bay Area who works on the issues of technology and civil liberties including privacy and free speech. I want to bring up at this point two comments.

The first more logistical, which is asking that the opportunity for public comment be better publicized and that the specific logistics for when, how be better explained in advance. So for instance, the details that were articulated at the beginning of this session were really helpful, and it would be great to have been able to read them in advance in the agenda.

And the second point I'd like to just put on the record is -- gets back to the stakeholder idea. The

policy you produce will touch on other domains, including innovation and expression more generally. And I want to make sure that you hear from experts who may not even think of themselves as privacy experts but people who may be experts in other areas or just practitioners or people who need to live with the consequences of how your policy will interact with these -- with their needs and the reality of other policy considerations, and to make sure that that's solicited and a part of the picture that's solicited and incorporated before any regulations are promulgated. Thank you.

MS. URBAN: Thank you very much, Ms. Gellis.

Mr. Gourley, is there further public comment?

MR. GOURLEY: Yes. Mr. Kloczko, you now have permission to unmute yourself. Thank you.

MR. KLOCZKO: Hi. Good morning, everyone. This is Justin Kloczko from Consumer Watchdog. And we're particularly concerned about precise geolocation in cars. The drafters of the CPRA envisioned the law would address overreaching the auto industry. We believe manufacturers do not need to know our geolocations to operate. Having geolocation for emergency services, for example, does not mean they can take our safety hostage and then sell or share our data.

This is a real serious issue for people. Just some

concerns I'll highlight quickly. The federal government reported that thirteen of the leading automakers collect, use, and share our data in order to track and market products without any really substantial limitation. If you're in a newish car, it's capturing everything you're doing, such as location, speed, braking, your buying habits, your text messages, kind of your total identity.

And you know, we've learned car manufacturers are working with software companies to use this data to bring advertising right into the dashboard feeding directly, you know, apps such as Domino's or Starbucks. So they can better know when a person is likely to buy, say, a cup of coffee. Data miners like WeHo, you know, tout its mobility data of over ten million connected cars and they claim to see precise speed in which cars are traveling on 95 percent of U.S. roads.

And geolocation can really meddle with insurance premiums. We've learned that insurance companies are working with the state insurance commissioner to allow telemedics data to calculate our insurance premiums, which we believe will -- will redline insurance policies and lead to discrimination.

And you know, one of the biggest misconceptions is that technology is making driving safer. It hasn't. Deaths are at all-time highs, prompting the federal

government to act recently.

So the intent behind the CPRA was for greater consumer protection, not less. But we're looking forward to regulations that end this era of data monopoly. Today, consumer watchdog is publishing our report, connected cars and the threat to your privacy. And so we just wanted to say that and thank you to all the board members for your time and patience.

MS. URBAN: Thank you very much, Mr. Kloczko.

Mr. Gourley, is there further comment?

MR. GOURLEY: There are no other comments at this time.

MS. URBAN: Thank you, Mr. Gourley.

As is my practice, I will wait just a little while in case anyone is formulating a thought as can be the case. And if not, then we will move on to the next agenda item.

MR. GOURLEY: There's no one else at this time. Thank you.

MS. URBAN: Thank you very much, Mr. Gourley.

With that, we will move into the informational presentations for the day, which you can find under agenda item number 5, informational presentations continued, overview of risk assessments and consumer rights with regards to public -- excuse me, to automated

decision making. You can follow along on the program for the day. Again, please note that we will take some breaks.

I will introduce each speaker with a short biography and then they will present to us. I understand that speaker bios and any slide presentations speakers use will be available on the agency's website as soon as they can be processed. We are also -- we'll have a transcript as well as the recordings. So again, you'll have lots of opportunity to review the information if you would like in the future.

With that, I am very pleased to introduce our first speaker for the day, Dr. Safiya Umoja Noble, who will be discussing data processing and automated decision making and challenges and solutions there -- about that. Dr. Noble is an internet studies scholar and professor of gender studies and African American studies at the University of California Los Angeles. There at UCLA, she serves as the cofounder and codirector of the UCLA Center for Critical Internet Inquiry, known as C2I2.

She holds affiliations in the school of education and information studies and is a research associate at the Oxford Internet Institute at the University of Oxford where is a commissioner on the Oxford Commission on AI and good governance. Dr. Noble's academic research

focuses on the internet and its impact on society. Her work is both sociological and interdisciplinary, marking the way that digital media interacts with the issues of race, gender, culture, power, and technology.

Dr. Noble is the author of Algorithms of Oppression, How Search Engines Reinforce Racism, which is not only an important academic contribution, but also a best seller. She's also the coeditor of multiple edited volumes.

Dr. Noble has won a number of prizes and recognitions for her groundbreaking work. I want to highlight a very special one, particularly for the public who don't necessarily -- to academics this is an extremely important award. In 2021, she was recognized as a MacArthur Foundation Fellow. These are commonly referred to as genius awards. They are given to people whose work has been truly groundbreaking and was given to Dr. Noble for her groundbreaking work on algorithmic discrimination.

MacArthur Fellows receive prize money to use as they see fit given the groundbreaking nature of how they think and the work that they do. Dr. Noble has founded a non-profit, Equity Engine, to accelerate investment in companies, education, and networks driven by women of color.

Dr. Noble holds a PhD and an MS in library and

information science from the University of Illinois at Urbana Champaign. And a BA in sociology from California State University Fresno. She has been recognized as a distinguished alumna by multiple of her institutions.

Dr. Noble is a board member of the Cyber Civil Rights Initiative, serving those vulnerable to online harassment. And was recently appointed a board member for the Joint Center for Political and Economic Studies, often thought of as America's black think tank.

Professor Noble, thank you very much for being with us today, and I will turn it over to you.

DR. NOBLE: Thank you so much, Ms. Urban. I am really pleased and honored to participate in this conversation and in this educational session.

I want to thank our brilliant team at the UCLA Center for Critical Internet Inquiry, particularly Akina, who is our policy director and is major contributor to preparing this presentation. And I want to say that some of what I may say today may be obvious for some members of the committee based on your own expertise. And some of it may not.

We thought it would be important as a public institution at UCLA to share our expertise and research that comes both our of our center and out of the field that is directly relevant to these processes and of

course there's much more so I would offer that the University of California system has many brilliant scholars and thinkers who should also be called upon in these processes and in the coming months and weeks to help share expertise.

So one of the things I want to say just briefly that brought me to this work is that about ten years ago or so, I started looking at -- well, I was leaving the advertising industry where I'd spent about a fifteen-year career in advertising right at the time that search engine optimization and kind of the ad tech business was starting to really take hold.

And I was thinking about the way in which systems like search engines and other types of digital media platforms were being relied upon by the public for deep information needs. People were using search technology in particular not like an advertising platform entirely, but also using it let's say in place of or in lieu of what libraries had previously provided for society or schools or teachers or professors or other kinds of subject matter experts, even parents.

And what I'd found as I was doing kind of a careful study of what happens in these advertising platforms is that there was a tremendous amount of misinformation, misrepresentative information. We now ten years later

have words like disinformation to describe the kinds of phenomena that we see on the internet when it comes to things that are patently false, things that are, you know -- take the shape of propaganda, if you will.

And I was really concerned with the way in which this would have disproportionate harm on protective classes, people who are already historically marginalized and have been historically and contemporarily discriminated against. And that really is kind of the impetus for the work that I've been doing for the past decades.

So what we find now ten years later is that while the conversation used to be that algorithms are for artificial intelligence, couldn't discriminate or couldn't be implicated in social harm in these ways because they were just math, now we have whole fields of digital studies and internet studies, and we have centers all over the world that are looking at this intersection between the internet and society.

And we understand of course that reducing algorithms and data and artificial intelligence and the whole ecosystem that goes into automated decision-making systems, reducing that just down to a concept like it's just math is a little bit like -- I think of it a bit like saying to biologist what is it to be human, and they

tell us that well, we're just cells and mitochondria. Obviously, that is true.

But we are also so much more than that. And part of that has to do with the social context within which our cells and mitochondria are interacting. And this is of course true about the way in which artificial intelligence, algorithms, AI-driven systems are social phenomena as well.

So let me just recap quickly some of the things that I think that you learned yesterday that kind of set us up for this brief conversation this morning. Yesterday you learned about the California Privacy Rights Act as a whole and its focus on automated decision systems, data collection, and social scoring. You also learned about the new consumer rights it affords Californians.

You learned about the role in rulemaking for how to hold companies accountable for protecting these rights. And this includes rules on how companies should be performing audits and assessments and what meaning information they must provide to consumers about the decision systems they're using.

So this morning, I'd like to address some of the harms that should also be addressed as you're developing your work. First of all, California is headed in the right direction with a truly robust privacy act that puts

the consumer first. As you can see, many place across the country have been trying to enact similar consumer rights focused legislation, specifically in relation to our data bodies or our data profiles or the data that describes who we are and is used to make decisions about consumers or the public.

The CPRA's focus on consumer rights in relation to data and tech also mirrors much of the consentful tech framework which we support. In the consentful tech framework, true digital consent is only achieved when its freely given, reversible, informed, enthusiastic, and specific. And so I'm going to talk a little bit about that.

Now in determining the new rules for enacting the California Privacy Act, we think you have a chance to do something groundbreaking. You can put Californians first in order to help you do this. And I want to offer some groundbreaking frameworks and suggestions that might support the way that you think about this.

First, I want to disvalue of any belief that tech, algorithms, or data are neutral. They simply are not. They're human-made and they reflect our society. Therefore, the policies that govern them, the rules you're writing, cannot be neutral either.

From that premise, the opportunities of the CPRA

floats the following questions. Since tech is not neutral, how do we define the meaningful information that companies much share about their tech to help people see and assert their rights per the CPPA mandate.

Since tech is not neutral, we also want to ask how deep must the CPPA rules require that rights for people and responsibilities for companies -- kind of how deep these rights and responsibilities go in order to support the people and moving towards a more just and equitable world. And for us, we really think about the social, kind of political, and economic context of justice and how justice and equity are realized or subordinated or subverted through a variety of different kinds of data practices.

Since the tech is not neutral, we also want to ask, how can automated decision systems be used to move from perpetuating the status quo to cultivating more equity and justice. And this of course has been I think a tremendously understudied area. Of course we are concerned with this in our center at UCLA but this whole framework if we're thinking about algorithmic discrimination, there are a number of organizations and people and we would be happy to also point you toward those people, to really help us understand the way in which data can be implicated in discrimination.

All right. So let's begin with this first question that since tech is not neutral, how can we define meaningful information that companies must share about their tech to help people see and assert their rights per the CPPA mandate. I think a large part of your charge is requiring businesses' response to access requests to include meaningful information about the logic involved in automated decision systems.

This is a very important task. Reminding ourselves that tech is not neutral, we know that the rules to define meaningful information must be designed to give consumers true power, especially consumers from historically marginalized communities and federally protected classes.

The tech, the automated decision systems designed by companies are not neutral because they are fundamentally designed with this kind of mandate, if you will, that they promote the company's profit. And one of the things that we find most challenging in this domain is that the ethical tensions around whether a company discriminates, whether it's implicated in designing technologies that are harmful in society is always put up against is the profit imperative. And even in some cases, let's say mandates that shareholder values be maximized at all costs. So those tensions have to be I think acknowledged

and in the process as you're moving through this work.

Sometimes this needs a product that is good consumers -- for consumers but unfortunately it can also mean that products that are good for some consumers are kind of predatory for others. So this is one of the cases where oftentimes tech companies are designing kind of for universal user and they have good results let's say a majority of users, but then there are outliers and vulnerable communities who are -- for whom those technologies might be weaponized or used in harmful ways against them or in predatory ways.

But you see this, for example, with advertisers of predatory financial instruments who use algorithms to target their ads to people who are the most in need of quick financial support and who are often targeted with a high interest loan. And they're also more likely, these consumers, to get trapped in the high-payback interest rate. And kind of a vicious cycle there.

So it can mean that a product that does well enough with consumers, you know, is strong in some ways but it doesn't necessarily account for people who are in the margins. And what we see is this often leaves historically marginalized groups excluded.

And another example I would give of a way that kind of technology works for the majority and then is

exclusionary for other would be technology that's used to screen and diagnose skin cancer. This often works well for people with lighter skin. But it doesn't work for people with darker skin. And so there are many, many different kinds of examples. I won't give you an exhaustive list today, but I will say these are the kinds of things that would be seemingly benign to a company and really we don't even know about the harm until consumers themselves or consumer interest groups or researchers see the kind of disparate impact.

All right. So from all of this, we often find that vulnerable communities fair worse in the design process of a variety of different kinds of technologies. And to define meaningful information in a superficial way leaves too much room for companies to give information that really reinforces their profits or speaks to the majority of the consumers for whom their product is being used.

It gives very little power to consumers, especially from historically marginalized communities to point out the flaws and the harm. So instead, meaningful information should present us with the opportunity to make public policy, to borrow a phrase from the Center for Urban Pedagogy. And in order to do this, meaningful information given by companies about their automated systems must include a couple of the following things.

First, information on the model used to make the decision. This must be fully and accurately explained without much abstraction in a way that is understandable to the people who are affected by the decision system. If it's too complicated to be explained to an affected person, perhaps companies should not be allowed to use that kind of decision model. If we could not explain the decision model to the people that are affected by it, we risk allowing decision that are kind of deeply impacting real humans to be made by models and machines that feel deeply abstract and nonhuman. So we kind of need to really address this tension.

We also think that information that is shared with the public should not be -- not only be shared about the model, it must show the impact of the model. By showing the impact on the individual, you know, we really help explain the -- and reveal the differential impact across different kinds of group or different kinds of people.

So we need to understand how these models are impacting individuals, but that alone is not enough. We really need to show the impact at the level of community too. So let me say a little bit more about this.

Sharing with an individual meaningful information that includes the model and how it determined its decision about that one individual reinforces assumptions

about the individual's undeservingness. And we know that this -- we see this over and over. It divorces, let's say, the individual decision from a community level understanding and really often divorces it from the historical context of how that decision was arrived at. It really reinforces the idea that there's a neutral or meritocratic process at play, and of course we know that this doesn't exist.

It isolates the individual and prevents any real accountability to communities. It's hard to prove, for example, that you were individually discriminated against by an algorithm until you start to see that more than 50 percent of African Americans were discriminated against. So you have to really understand yourself in your own community class if you're part of a protected class, whether that decision model is affecting the entire class that you are a part of.

If you can't see how the decision about you as an individual is tied to other groups, then it's very difficult to get at how the discrimination is happening. And what we often find is that people just feel that they, you know -- the bank didn't approve them for the loan and they -- it's only until investigative journalists reveal that, let's say, a bank -- I can think of one -- that was just in the headlines two weeks ago.

That a bank denied loans to 50 percent of the protective class that they're a part of. And then we start to understand that maybe there is a problem here with the data model.

All right. Let's take another example where meaningful information has failed because it's centered on the individual and it's vague and where it's been improved upon at least by adding this community level analysis.

So the image here that you see on the left is from one of the researchers in our center and it's their interest category profile on Facebook. The interest categories are an example of Facebook's attempt to give what they might call meaningful information to users. Facebook provides each user a list of their interest categories, categories that advertisers can use to refine the audience for their ads.

Upon first look, it seems pretty benign. And it seems like Facebook has indeed done away with the racial profiling categories that they were sued for in 2018 by the National Fair Housing Alliance and promised to take down. But when The Markup, investigative journalism news outlet, compiled community level data using their project Citizen Browser, it showed a different story. Citizen Browser is a panel of 1,300 paid participants who

provided The Markup with their demographic information and allowed for periodic capture of their Facebook feed data. Data that is shared to The Markup where they can make community level connections rather than seeing the isolated experiences of an individual that can be explained away.

Their research showed that these interest categories served as proxies for race categories, allowing Facebook to skirt their commitments to taking down racial profiling options for advertisers. Seeing an individual's interest categories alone, you might not immediately realize the relationship it has to bigger communal experiences.

This person in this profile I'm showing you who is black has in their interest category Black Lives Matter. It's an interest category that they didn't choose but the Facebook algorithm assigned them. The Markup's research shows that this was one of the categories that advertisers could use as a proxy to filter their ads to target or to exclude black users.

Facebook's individualized interest categories don't even show us the model that it's being used. And Citizen Browser doesn't have the insider information to show us the model Facebook uses either. It's clear that Citizen Browser is an improvement on meaningful information since

it shows the communal impact, not just the individual impact.

Now imagine if Facebook had to show that communal impact as well as provide further detail on the model that it used. Consumers could truly assert their rights and demand products that are designed to benefit them, not just company profits. And of course they also might be able to see the ways in which they are targeted with predatory products or steered toward particular kinds of ideas, including propaganda that often circulates in Facebook.

So let's return to the second question. Since tech is not neutral, how deep must the CPPA rules require -- excuse me, my long-term COVID effects. I apologize here. How deep must the CPPA rules require that rights for people and responsibilities for companies go in order to support the people in moving towards a more just and equitable world.

So since we can establish through many different types of research and books that the technology itself is not neutral, shallow rights and responsibilities will lead to superficial attempts to satisfy consumer rights in order to protect company profits.

Superficial rights and responsibilities will never address the deeper societal problems that tech replicates

by default. So rights and responsibilities must go as deep as possible up and through the supply chain. This means that rights like digital amnesty, reversible consent, right to know, right to delete, the things you've reviewed yesterday or a year ago, must happen all the way through the supply chain.

Audits and assessment are one step closer to holding the companies and their technologies accountable, but they really are not enough. Many reports have shown the limitations of audits and assessments because companies refuse to give the information, governments refuse to define what automated systems are or the audits and assessments have insufficient penalties.

To overcome this limitation, more rules should be made to require companies to delete algorithms and the data associated with these algorithms when these audits and assessments fail. Deletion of algorithms and associated data as the ultimate recourse is necessary because it helps us shift away from the logic of if only we fixed or debiased the algorithm.

And this of course is a very prevalent argument that you will hear -- you probably have already heard -- which is this idea that we can somehow debias the algorithms or that we can kind of fix this at the level of -- by just kind of tweaking the tech. What this logic focuses on is

validating the product, not on supporting the people who are impacted by the product.

So if we commit to supporting people's rights and businesses' responsibilities through the supply chain throughout the design process throughout the development process so deeply that we're willing to tell a company, that if they continue to fail, they must delete the algorithm and the data, then we're truly listening to consumers and not the companies' kind of profit motivations and the examples that they will give that will kind of justify and support their own profit motives.

An example of where this logic of deep rights and responsibilities has failed with devastating consequences is the California gang database. CalGang, a statewide gang database developed in the 1990s, has had hundreds of thousands of Californians listed as being members. At its height in 2012, CalGang had over 200,000 people listed as gang affiliated.

In 2016, a state audit of CalGang showed that among those listed in the gang database, 42 entries had birthdates that indicated that they were one year old or younger, with the majority of those entries being in the database because the individual had allegedly admitted to being a gang member.

The 2016 audit also showed that despite federal legislation that mandated that people would be removed from the database after five years unless updated with subsequent criteria, auditors found over 800 individuals who should have qualified for being purged but were still on the list. Almost half of these individuals had purge dates set more than 100 years into the future.

Individuals can request to be removed from CalGang. But according to Urban Peace Institute's 2018 report and first-hand experience supporting dozens of removal requests, the process is ineffective. Removal requests are often denied, people don't know they can make such requests, and it is unclear if removal protects from erroneously being added to the database again in the future.

Oversight and rulemaking related to CalGang has undergone consistent legislative updates since at least 2013 and it continues to fail. Starting January 1, 2020, the Department of Justice will now do regular audits of CalGang. But at some point, you have to ask, when is enough enough?

Individuals trapped in this work in technology suffer consequences that go beyond the technology itself. The CalGang database is used for employment and military related screenings. A mother responds to police officers

at her front door wanting to question her six-year-old child about his gang affiliations. A 59-year-old man is added to the list after playing chess in the park with friends. The list goes on.

With technologies like gang databases, automated decision like decisions about where to send police is rooted in the faulty human process of adding names to the database in a policing system that is rife with individual racial animus and systemic racist policies and practices. Audits and assessments keep giving a pass to CalGang to fix itself, to debias itself. But these efforts are not and will never be enough.

CalGang has had enough time to improve since its implementation in the 1990s but it continues to fail Californians. Imagine a world where after a few failed attempts to fix or debias CalGang, that it was scrapped instead of being able to stick around and continue to cause harm in real people's lives.

Now, for our last question. Since tech is not neutral, how can automated decision systems, technology, algorithms, and data be used to move from perpetuating that status quo to cultivating equity and justice?

While this question might not be the exact purview on this board, so much of the tech you will see, assess, audit, and engage with will be tech that perpetuates the

status quo.

In this case, when we say tech that is not neutral, it means that tech not only promotes company profits over consumer rights, but that tech even with the most default and passive decisions reinforces the status quo, a status quo that privileges some and punishes others.

So what we would need -- what would we need to overcome this? What would be the technologies we would be excited to see? What is needed for an ADS to promote equity and justice?

The purpose of the automated decision system must be to address and redress historical structural racism. Like the popularized framework of being racist or antiracist, there is no neutral ground in ADS creation. So to promote ADS, ADS that support equity and justice, requires proactive, pro-equity, pro-justice design. You must consider who benefits and who is harmed by the ADS and its designed process.

If you don't, you'll fail to address and redress the historical structural racism, sexism, other kinds of class-based discrimination that is often baked into the data sets that are used to train machine learning algorithms and kind of help steer and guide automated decision making systems. And that means you'll likely recreate it.

And lastly, I want to assure you that pro-equity, pro-justice race aware algorithms can and do exist. They were unfortunately hard for me to find for this presentation, a reminder that there are too few and far between, but for the lawyers here today, I want to emphasize that algorithms can be race aware without triggering disparate impact and affirmative action legal logics.

In fact, I strongly caution us away from the assumption that any algorithm that -- algorithm that considers race and justice must by default be defined by affirmative action logic. Such an assumption reinforces the false notion that historic privileges are married to -- kind of part of the meritocracy that's gained and that the status quo is fair.

All that to say there is a tremendous amount of work that's being done right now to think about the way in which algorithms and automated decision making systems are perpetuating historical discrimination and there are people working on logics that would help reframe and help us think about true equity in society.

So let's think about this again in the real world. The examples that I've laid out already explored the failure and potential for automated decision making systems to be used to further equity and justice.

CalGang, as I mentioned, and the decisions that flow from it, police visits to houses, employment screenings, deployment of police, and kind of disproportionately into poor black and Latinx communities, southeast Asian communities, the question is does it address and redress historical racism.

The obvious answer is no. It perpetuates the historic racist policing practices of our country and it makes marginalized communities even more vulnerable. It leads to more surveillance and stops by the police, making both police and the people feel like they are criminal by default, you know, this kind of orientation to being deployed out to certain neighborhoods reinforces this idea of inherent criminality of those communities.

And of course, who does it benefit and who does it harm are always the questions that we're asking. In general, it benefits the CSRA International Inc., a for-profit company that designed the CalGang software. It harms the mother who had the deal with the police officers who came to talk to her about her six-year-old son. It harms the 59-year-old man who wanted to play chess with his friends and instead the police were deployed to the park.

This is a technological system that does not cultivate equity and justice. It creates lists and

automated decisions that reinforce inequality and racism. And there is a company that truly profits from that. On the other hand is an example of a technology that gets much closer to cultivating equity and justice. Clear My Record, a partnership between Code for America and the San Francisco District Attorney's Office, works on automated record expungement.

The underlying law, California's AB1793 is a law that tries to address historic racism. AB1793 passed in 2018, mandates that counties' clear convictions for buying or possessing marijuana, convictions that disproportionately were targeted towards communities of color.

The law itself aims to decrease police surveillance and the impact of criminal records on the very communities that have felt the brunt of racist police surveillance, a racist criminal legal system, and a racist public policy that was deployed through the war on drugs. Adding an automated expungement technology on top of a law that is moving in the direction of addressing historic racism means the technology can support that good underlying law.

With Clear My Record, the tech is dependent on the good privacy -- of the good policy of AB1793, versus with CalGang, the tech defines and reinforces the regressive

policy and the policing system.

So these key questions are helping us determine if technology or an automated decision system is cultivating equity and justice, and these are the same questions we can be asking in designing our policy. And it's through these questions that I formulated by recommendations for defining meaningful information and how deep rights and responsibilities should go.

As the CPPA board continues to define their rulemaking in this process, I urge you to use these grounding questions as you define your rules. Thank you so much for this opportunity to share these ideas with you and I'm happy to take any questions.

MS. URBAN: Thank you very much, Professor Noble for that incredibly helpful presentation. We thank you very much for your time and for sharing your expertise with us.

I'm delighted now to introduce our next speaker, Dr. Gwendal LeGrand, who will be presenting on data privacy impact assessments, what should be considered. Dr. -- excuse me, Dr. LeGrand is the head of activity for enforcement support and coordination at the European Data Protection Board. He is particularly involved in the coordinated enforcement framework and the support pool of experts, which aims to assist the national supervisory

authorities in their investigations and enforcement activities of significant common interest. Before joining the European Data Protection Board, Dr. LeGrand worked at the French data protection authority, the CNIL. I'm afraid if I try to say this in French no one will understand me, so I'm going to say it in English, the National Commission for Computing and Liberties, where he was deputy secretary general from 2019 to 2021, director of technology and innovation from 2014 to 2019, and head of IT experts department from 2007 to 2014.

At the CNIL, Dr. LeGrand focused on new technologies, information security, digital transformation, ethics, and international affairs. Dr. LeGrand served as the coordinator of the technology subgroup of the European Data Protection Board from 2018 to 2021 and he served as the board's liaison to ISO/IEC JTC 1, SC 27, WG 5 working group, which developed privacy standards and is a member of the advisory group of ENISA representing working party 29 and EDPB since 2015, excuse me, of the European Data Protection Board, since 2015.

Dr. LeGrand started his current academia and was an associate professor in networking and security. He received his PhD in computer science from the University of Paris 6 in July of 2001, and his master's degree in digital communications from Telecom Paris in 1998. He

graduated as an engineer in telecommunications from SudParis in 1997.

I am very pleased and grateful that you are here today, Dr. LeGrand, and I will turn it over to you.

DR. LEGRAND: Thank you. Thank you very much for the introduction. I hope you can see my screen and my slides. So today -- it's okay? Yes?

MS. URBAN: Yes.

DR. LEGRAND: Today I'm going to talk about data protection impact assessments and GDPR which we call DPIA, it's currently called PIA, privacy impact assessment in many regions of the world. And as you know, we have this concept that was introduced in the GDPR in 2018.

So in Europe as you know, GDPR has replaced the data protection directive that had been adopted in 1995. And a directive is a minimal harmonization law. It's a level that has to be transposed in member state law. In the directive, there was this provision on private checking for risky processing. And basically provided that for risky processing, some prior checking had to be done by data protection authorities.

So to a certain extent and without being overly simplistic, I would say that compared to this directive and to the associated national law, what GDPR has done is

that it has shifted the prior checking by the authority to a kind of accountability obligation to carry out DPIA when the processing is likely to result in a high risk. I really emphasize the word "likely" because the objective of the DPIA is to ensure that once you have implemented the appropriate measures, the processing is not high risk anymore.

And in GDPR, you also see that once you have implemented measures to mitigate the risks, if the residual risk is still high after having performed your DPIA, then you're obliged to consult with a data protection authority. In practice, this should not happen, because the objective of DPIA, as I said, is to find the appropriate safeguards to mitigate the risk that you identify through that process.

So DPIA is a formal process which you find you can -- the relevant material is Article 45 of GDPR and Article 46 of GDPR. It's not box ticking exercise and if you want to know more about GDPR -- about -- sorry, about DPIA, you can also consult the EDPB guidelines on data protection impact assessments, which the link is on this slide here. And there is a lot of material also that was done by some national authorities and I will present this during my presentation.

So today I am going to focus mainly on the cases

when an organization is required to perform a DPIA. And I'm going to explain how we do a DPIA and basically walk you through the relevant material that can help you in conducting a DPIA and conducting this exercise.

As I said before, the GDPR requires that you as an organization have to do a DPIA when the processing is likely to result in a high risk. GDPR gives you free examples when processing is likely to result in a high risk. In a nutshell, it is when you have automated processing which is likely to result in decisions that produce legal effects. Second example is large-scale syndicated data. And third example is monitoring of a public accessible area on a large scale.

GDPR also says that the authorities must further specify the cases when you need to do a DPIA. I'll come back to this in the next slide. And they can also specify lists of cases of which a DPIA is not and never required. And I will also present this in the next stage.

The objective of the DPIA in the GDPR is really to build and demonstrate compliance. So you need to do an in-depth analysis of your processing operations. You describe them, you understand and you describe necessity and proportionality of the processing. You identify the risks and you identify the measures to mitigate the

risks. So it's really a risk assessment exercise applied to privacy and data protection.

It's something you do prior to the processing. And it's often compulsory, meaning that if you haven't done a DPIA when this was required, GDPR provides for some sanctions. It can go up to 10 million euros or 2 percent of the annual turnover of a company. And GDPR uses always the highest number that will be found for the calculation of the fine.

Now, the interesting question is when do we do a DPIA, because to a certain extent GDPR is not extremely helpful when it says only that you need to do it when the processing is likely to result in a high risk. So there's a group of European authorities, which used to be called the Article 29 working party and now is called the European Data Protection Board, which is a body of the European Union that was created by GDPR, and one of the missions of this group is to give some guidance on how to implement GDPR correctly. Guidelines were adopted -- were prepared and adopted by Article 29 and were endorsed by EDPB; the link is on the previous slide that I showed you before. And one of the things that we did when we elaborated the guidelines was to try to give some guidance on when to perform a DPIA. We did a bottom-up exercise, actually, to do this, so we went to

all the authorities and we said, well, give me a list of cases when you think a DPIA should be required, and we ended up with a list of 100 to 150 processing types of processing operations, which was not very practical and workable. So what we tried to do is to group them together on the basis of different criteria, and this criteria are the ones that you find in the guidelines and that are presented in the slide here, on the right side of the slide. We've identified nine criteria. I won't list them all, but you can see them on the slide, so evaluations/scoring, systematic monitoring, large scale, and so on and so forth, and what we have seen empirically is that whenever two out of these nine criteria are met, there's a strong recommendation to perform a DPIA. So it's a rule of thumb. In some cases, if there's more than two or -- if there's two or more than two, it means you need to think twice if you decide not to do DPIA. And in some cases, if there's only one of the criteria that are met, it's recommendation is a good practice to do DPIOs.

These nine criteria have been also transposed to a certain extent in national documents that were adopted by national authorities, because in the EU, you know we have the European level and we have GDPR, which is the European law. And then in each of the member states,

there is a national law that complements certain aspects of GDPR, and there's a data protection authority, which is an independent authority, that needs to adopt certain documents and is in charge of enforcement. One of the requirements of GDPR, as I said before, is for these authorities to adopt lists of cases for which a DPIA is required.

So what they've done is they've taken these criteria. Most of the time, they produce the lists. They sent this list for an opinion to DDPB and DDPB checked whether or not the items that were included on the list were making sense and were harmonized with what was done in the other member states. So EDPB adopted opinions, sent this back to the national authorities, and on that basis they published each of them -- in each member state, the authorities published their national list.

In a nutshell and if you want to have the helicopter view and on the stand when to do the DPIA, I think that the nine criteria that you have in the guidelines are very good guidance on which to rely on to -- and when two of these criteria are met, it's interesting for a controller to think twice before implementing its processing operations and make sure that they have the appropriate safeguards implemented in the system.

There's another list, which is optional in GDPR, which is the list of processing operations that are exempt from doing a DPIA. Now, this is interesting because it can -- you can either think about including in this processings that will never be high risk, of course, but you can also have cases where it's a processing operation that is very generic and for which the exercise of conducting the DPIA has been done already and if you implement the processing operation in the way that is described in the framework and using the safeguards that have been identified in this kind of generic DPIA, then you can be assured that the processing will not be high risk. So it's really -- we really have these two types of lists in the law. The compulsory list is required from the national authorities, and the exemption list is optional.

On this slide I'm trying to show you all the documents that can be relevant for you when you think about exercises that are similar to DPIA's as described in GDPR. So there's GDPR, of course, which I mentioned before, article 35 and article 36; there's EDPB guidelines that I have mentioned already. If you go to certain member states, some member states' authorities have also issued some guidance, and as explained in the introduction before, I worked many years at the CNIL,

which is the Commission Nationale de l'Informatique et des Libertés, the French Data Protection Authority, and we have -- we had there long history of working on privacy risk management. Back in 2012, we had done some privacy risk management guides that were published in French and in English, and they were revised and enhanced in 2018 to match the requirements of GDPR when GDPR became applicable.

So you have three guides there that you can find on the CNIL websites. I don't know if you can see -- this is the webpage of the CNIL, but there's a page that is dedicated to PIA. There's a description of the elements I mentioned before, so when to do DPIA and so on and so forth, and what you will see in the three guides is how to do DPIA. So there's a methodology that is described in the first guide, which is aligned and compatible with the risk assessment exercises that people know quite well and are acquainted with in the information security world. So in a way, what was done there is to transpose risk assessment for information security to the world of privacy. So thinking about not the impacts for the organization, but thinking about the impacts for the data subjects for the individuals. So this is what you have in the first guide, which is available in English. The second guide you will have a list of templates, which

explain -- it's a kind of framework to conduct the DPIA. And the last guide is a list of measures that you can implement in the system to mitigate the risks that you identify with the methodology.

What was done on top of that was to publish a number of use cases, so what we sometimes call the PIA frameworks, and we've done one for internet of things, for connected objects for instance. That is also available on the CNIL website, and this PIA framework basically is guidance provided by the authority that gives you the list of typical questions and typical answers that can be relevant in the specific sector.

One other thing that was done by the authority, and I'll come back to this in more detail, is to publish some software, because -- I used to say that it's very nice to do some guidance to publish some PDF, but if you print all the documents, this represents more or less 200 pages, so it's very lengthy. It's not necessarily easy for people who are not used to risk assessment exercises and risk management exercises to know where to start, and therefore there was an exercise that was done at the CNIL to edit some software and publish some software, which you will find on the CNIL's website, and I'll come back to this on the next slide.

Last thing that was done was to work at ISO. ISO is

the International Sanitization Organization, and there's a working group there that is developing some standards in the field of privacy. And one of the standards that was developed is 29134, which is a standard on privacy impact assessment. So we -- at the CNIL, we were participating in the work of ISO, of the working group that ISO that is drafting those privacy standards, and we contributed to the 29134 to make sure that what is in the method and what is in the software that I mentioned before is compatible with the international standard, 29134, that is now recognized at international level.

Just a few words on how to conduct a DPIA and what has to be included in the DPIA. This is explained in GDPR in article 35.7 and basically is what I said before; you have four parts in DPIA. I've described the processing, I evaluate necessity and proportionally of a processing, I identify the risks, and then I mitigate the risks. And so it's really accountability, an accountability process that is quite familiar to people working in the area of information security. This is the same thing with a bit more detail, and this is what is described in much more detail in the guides and in the software that I've been talking about just before. Describing the context, which was the first step of the DPIA, means you describe the processing, you describe the

type of data that are going to be processed, you describe the supporting assets, so where your data is stored and how and so on and so forth. So it's a systematic description of the processing.

And there -- this again is a bit over simplistic, but it's just for you to understand the process. There's a legal assessment by lawyers; this is the last part of the slide. So you list the measures to protect the rights of the individual, and you check that this is in line with the requirements of GDPR, and we do this assessment of necessity and proportionality. And on the right side I've put technologists. You do this cyber security assessment exercise. So you check which controls you have implemented in the system and you check whether or not they're sufficient to make sure that your processing is not high risk anymore at the end.

And then there's a decision to be made by the organization. You produce a report and you assess whether or not the risks are going to be acceptable, whether you took them down to a level that is not high risk anymore. If it's not the case, you reiterate the process, you include more controls, and if it's impossible to go down to processing that is not high risk anymore, GDPR says you need to consult with the authority that will give you some guidance or tell you not to

implement the processing operation.

A few words now on the PIA software that was edited by CNIL. So this was a decision we made back in, I think, 2016 or 2017 to try to explain how to do DPIAs and help especially small organizations with DPIAs, because big organizations have CSOs, they have people who know this type of process, but small -- for small organizations, this can be a huge burden to do DPIAs. So this is the reason why there was this decision to make this tool at the CNIL and make it available.

This tool is software that you can download that is standalone on your computer or it can run on the software. It's -- it was initially released in two languages, French and English only, but it was published on GitHub, it's open source, and now it's available in more than 20 languages because people throughout the world found it interesting and contributed their language version. And basically what this tool is doing is that it walks you through the different processes that I've been describing before.

So you'll see it very briefly here, perhaps not very clear, but you have screens like the ones that are presented here on the right part of the slide, and it takes you step by step through the different steps of the PIA. You fill in the sections, you explain which

controls are in place, and on the right part of the screen, you have some information that is contextual, depending on the part that you are filling in to help you fill in the PIA and based on material that is contained in the guides that I was describing at the beginning of the presentation.

So it's work that was done also with a designer that was hired at the CNIL to try to make this as user friendly as possible, and the outcome basically is you have a software that can collect all different PIAs that you are conducting. It does the risk mapping, so explaining which are the high risks at the end of the risk assessment exercise. It represents the risks in the form of a map, and you can identify the controls to mitigate the risks and see how to take the risks down to an acceptable level. I forgot to mention that the risks are always described according to two dimensions, likelihood and severity. And again, in the guides you will find some hints on how to do this risk assessment and how to quantify the risk in terms of severity and likelihood.

At the end of the DPIA, what you do is this risk assessment -- risk, sorry, acceptance decision, and once the organization decides that the risks are mitigated in an appropriate way, you can proceed with the processing.

One recommendation, which is not a requirement of GDPR, is to publish the PIA or publish a part of the DPIA, also to show to the public that the risks have been tackled in an appropriate way and that the exercise has been conducted seriously by the organization.

So with this quick summary of my presentation, the PIA or DPIA in GDPR is an exercise that makes it possible for you to build and demonstrate compliance of processing operations. It's something that feeds into more general processes that can be implemented by companies like, you know, audits, the register of processing operations, risk management exercises, management of information security, and so on and so forth, and the interesting part is that organizations are quite familiar today with information security management, risk management, and privacy risk management, or data protection impact assessments, or exercises that are extremely similar to this. It's really the same logic; it's just that the focus is not only on dealing with and managing the risks for the organization, but also focusing on the risk for the individual.

Thank you very much for your attention. I just highlight that the slides that I showed you today, most of them have been adapted from material that was developed previously when I was at the CNIL, and this is

why there is this CNIL credit. Thank you very much, and I'm ready to answer your questions.

MS. URBAN: Gwendal LeGrand, thank you very much for that very helpful and relevant presentation. It's greatly appreciated. Thanks again to actually both of our first two speakers.

We are now going to take our first break, which is on the longer side, to accommodate our speakers' schedules. We'll reconvene at 11:30 a.m. Pacific Time for our next presentations. Please feel free to leave your video or teleconference open or to log back -- log out now and just log back in at 11:30, and we will look forward to seeing you then. Thank you.

(Whereupon, a recess was held)

MS. URBAN: Thank you. Mr. Gourley, are we ready to return to the meeting?

MR. GOURLEY: Yes, Chairperson Urban, we are ready.

MS. URBAN: Thank you very much, Mr. Gourley.

Welcome back, everyone, to the California Privacy Protection Agency's March 2022 pre-rulemaking informational sessions. I would like to remind everyone that we are recording this meeting. If you're just joining us, we are listening to a series of presentations under agenda item number 5, Informational Presentations Continued: Overview of Risk Assessments and Consumer

Rights with Regards to Automated Decision-making. We have three more presentations today, and then we will finish the day with public comment. Excuse me. my apologies. I'll remind everyone of how to engage in public comment when we get to that part of the day, and please note, we may also take a short break at some point. It won't be probably as long as the break that we just had.

All right. We will now continue on with our first set of informational presentations. If you'd like to note the place on the agenda, we are on item -- agenda item number 5, part c, cyber security audits. And if Professor Hoofnagle is ready --

Good morning, Professor Hoofnagle.

I'm delighted to introduce our speaker on the topic of cyber security audits, Professor Chris J. Hoofnagle of the University of California-Berkeley. Professor Hoofnagle is professor of law and residence at the University of California Berkeley's School of Law, where he teaches cyber security, programming for lawyers, and torts. He is affiliated faculty with the Simons Institute for the Theory of Computing, a professor of practice in the school of information, and the faculty director of the Center for Long-Term Cyber Security. An elected member of the American Law Institute, Professor

Hoofnagle is of counsel to Gunderson Dettmer, LLP, a firm in Silicon Valley and serves on boards for Constella Intelligence and Palantir Technologies. Professor Hoofnagle is a prolific and far-eyed author in the areas of privacy, cyber security, data protection, consumer rights, and emerging technologies. His recent books include In Law and Policy for the Quantum Age with Simson Garfinkel and Federal Trade Commission Privacy Law and Policy.

I know from having Professor Hoofnagle as a colleague at UC Berkeley that he is an extremely innovative thinker and a dedicated and innovative teacher. Professor Hoofnagle holds a BA and a JD from the University of Georgia, and we are delighted to have him here today.

Professor Hoofnagle, please take it away.

MR. HOOFNAGLE: Thank you, Professor Urban. I'm delighted to have this opportunity to present for the agency. I'm going to share my screen and make sure this is in order. So let me just ask: Do you have the full slide?

MS. URBAN: We do. We have the presenter view.

MR. HOOFNAGLE: Okay, let's try that.

MS. URBAN: Now we just have the full slide.

MR. HOOFNAGLE: Okay, great. Wonderful. So I'm

just working on my environment here. Let's see if that works. Okay, great. Thank you for having me today.

This high level presentation makes four points for you to consider as regulators in the security space. Some of you will already be familiar with these ideas, but I'm going to stay at a relatively high level given the time constraints and the complexity of security regulation. Remember security is a process. It's never completely achieved. Today I'm going to talk about four dynamics in that process, four dynamics that I think you will see as regulators in the security space.

The first is that familiarity with the terms of art used in security is important because security terms have counterintuitive meanings in practice. Second, I'm going to revisit the CCPA's terminal policy goals, and I do this to warn you that instrumental activity surrounding security can overshadow these goals. Third, I'm going to explain how security frameworks are highly congruent, meaning that at the highest level, there is consensus about what good security hygiene is nowadays. And then fourth and finally, I'm going to explain that there are many policy options for implementing security frameworks. Let me start with -- yeah, here we go -- terms of art in security, three confusing distinctions, and it's important that you're familiar with.

First, it's important to know that there's a difference between audits and assessments. Audits are examinations against an externally defined standard. These standards are often objective in the sense that they have a pass/fail basis. So for instance, if there are less than two millimeters of tread on your tires, you probably need new tires.

Assessments are different than audits in two critical ways. First, in an assessment, the client company gets to define what the goal is. The company can say, look at my tires, but even if they're worn, we have other ways of keeping the car safe, so I can still meet my goal even if my treads aren't more than 2 millimeters deep. Maybe the car is only operated in a warm climate, so it doesn't matter as much how deep the treads are.

The second difference is that in an assessment, the examiner is free to draw from a wide range of evidence to develop an opinion of compliance with that goal. Let me emphasize: It's an opinion; it's not a straight up or down. Perhaps the examiner, returning to the tire example, says you pass the tire test even though your tires are worn because you tend to drive slowly or because you never drive in the snow and rain.

Critically, most privacy and security evaluations are assessments, not audits. This means that the

independence and ethical standards of evaluators are key. At the Federal Trade Commission, for instance, the agency has started prescreening evaluators to ensure that they have a sufficient reputation in the field. If you ask a company for their security assessment, they may respond with a one-page-long opinion letter from that evaluator. So it's important to anticipate that outcome and not just ask for the assessment, but to ask for the underlying evidence that supports the opinion, the final opinion made by that evaluator.

There are also important differences between the term security incident and security breach, and apologies if this is too basic, but it -- there's an important policy point here. A security incident is any event that imperils confidentiality, integrity, or availability of information. For instance, if logs indicate that someone might have gained access to an account without authorization, that is a security incident.

Security breaches, on the other hand, are legal events. A security breach is the determination that an incident requires notice. So for instance, if that log, upon investigation, leads to a finding that covered information was accessed by some outside person, notice probably has to be given to the user. Network and software engineers will be the ones who identify and

diagnose security incidents, whereas the lawyers are in command of so-called security breaches.

What this means for you is that you will hear a different story about security if your evidence collection comes from the technical people than if it comes from the lawyers. The technical people are going to see more events, they're going to see more threats, and they're going to see events that did not lead to notice even when there's problems with confidentiality -- data confidentiality.

Finally, I want to mention the last term of art here. The word "accountable" has a strange meaning in privacy land. That is many in the industry use the word "accountability" to mean that they are able to make an accounting, as in we can tell you what happened to personal data. This is different than the use of accountability in everyday use, by which we typically mean that we are accountable for our actions through prosecution in the criminal justice system or through civil liability.

Let me turn quickly to this, the second point. A second major theme surrounds our terminal goals in the CCPA. You are familiar with these, and so I have just inserted them here on this slide. Let me emphasize that security is broadly defined in your policy goals. It's

the object of security is personal information, not just sensitive information and not just information that is somehow economically valuable, but rather all personal information. And the threat to be protected against is security incidents, not a narrower concept of security breach.

Why this matters, you heard a bit about this earlier today from Professor LeGrand, who talked about risks. The reason why it matters is that our concepts surrounding security can diverge. Your agency is charged with protecting a value, the right to privacy. Companies, however, will define this value through the lens of risk and use controls to manage that risk. In so doing, companies are likely to treat the controls they implement as the terminal goal rather than the policy aims of the CCPA.

I want to use the opportunity to talk a bit about policy goals and visions for security that could emerge from your agency. Security today is like the automobiles of the 1960s. We have powerful and awesome cars that gave us lots of utility in the 1960s, but they also ignored safety, and the industry blamed accidents on drivers. Since then, we have implemented technologies like the seatbelt to mitigate harm and advances like traction control that helps us avoid accidents in the

first place.

What would it take to move computer security along in the same way that makes markets and regulation drive improvements and automobile safety? One might be improvements in situational awareness. So here on this matrix here, I can tell you where we are today. We are in the known known category. We know that there are a fantastic number of security breaches. What we don't know is how many of these breaches go unreported, and no one knows about undiscovered incidents and security vulnerabilities out there. So we can think about changing the knowledge model of our security vision to move out of this category of knowing about breaches to knowing about other things.

Let me get my slide to advance here. We could also reconceive of the liability model for breaches. Currently, we follow something that resembles a negligence model, where agencies such as the Federal Trade Commission examine corporate practices and bring cases when those practices indicate lack of due care. This means that a lot of agency resources are tied up with the investigation and the determination of wrongfulness. It also means that there are many breaches where there's no relief for the consumer. One could think of those as non-negligent breaches, if you will.

Now consider an alternative model that looks more like enterprise liability. In this vision we are less concerned about specific wrongful practices and more concerned about making consumers whole. This model asks businesses that benefit from collecting information to assume all the downsides when those businesses lose personal information. The agency could also move towards a maturity model, where your security goals shift over time as the security landscape matures. In fact, you could become a driver of security maturity.

So the agency could start with a simple goal of getting all institutions to evaluate security risks. Believe it or not, some institutions don't. But as we look over the horizon, we could imagine pushing companies to deepen their examination of risk, and in the 2030s maybe we could imagine a solution where businesses are wholly owning the risks they create instead of externalizing them onto the public.

The third topic I wanted to brief you on is the amount of high level harmony in security frameworks themselves. The most popular security framework is the NIST cyber security framework. It was created to encourage security in the critical infrastructure sector; however, its flexibility and universality has made NIST attractive to many different sectors. Let me explain

what we're looking at here. At the highest level, the NIST cyber security framework identifies five key functions for cyber security. This is an example of the identify security function. Identify includes the concept that in order to secure an enterprise, one needs to know all about the systems that comprise it.

Moving on, you'll see I have some arrows here, so that's the function arrow there. Let me move onto my next slide here. However, the categories and subcategory columns decompose the high level identification function into more specific steps, but please keep in mind that even the most simple step here is not simple, it's complex for even small or mid-size companies. Just imagine in your own personal life if you had to follow this recommendation, ID-AM-1, and inventory all the physical devices and systems in your household, just imagine how long that would take and how complex it would be, and if you zoom out and imagine well what would that mean to also do that function for software and what would it mean to do that in a small- or medium-sized business, these are not simple tasks.

Finally, let me call attention to this last column. This is what's known as the informative references column. This last column presents a huge takeaway for businesses. There's a great congruence among security

standards. So this idea that one should inventory systems is shared amongst all the major security standards, and this column shows those cross references. That might seem like a pedantic point, but in fact it's very important for businesses.

Businesses may be subject to several or even dozens of different security frameworks, thus they want to build processes that will satisfy as many standards as possible. This offers a lesson for the agency. If the CPPA decides to adopt a new security framework, think carefully about that because creating a new one creates a lot of headaches for industry, and it might be largely duplicative of existing security standards out there. I think the thing -- the challenge you have to think about is whether you would materially advance cyber security by creating a new standard or whether our cost benefit analysis would reveal that going with existing standards is good enough and lowers compliance burdens.

Finally, let me conclude with my fourth point. While there is a high level consensus now about the steps one should take to promote good security hygiene, the mechanisms for implementing and assuring those steps are taken are very different. Companies like the NIST framework because it's voluntary and because a company can choose from a broad menu of precautions. The risk

here with the NIST framework is choosing poorly. Other approaches command specific compliance objectives. For instance, the opposite approach to NIST is found in PCI-DSS. PCI-DSS, which is the standard required of all businesses that handle payment card data, is highly prescriptive. It tells you exactly what to do and how to do it.

Now, somewhere in the middle exists other approaches, such as process- and control-based approaches. Process-based systems are similar to NIST because fundamentally the company is in command of defining what the goal posts are and how to meet them, but a controls-based approach is somewhere in the middle. In a controls-based approach, the company gets to explain how it will reach goals, but some outside entity defines what those goals are. So an agency like CPPA could articulate a series of goals and give companies the flexibility to choose how they reach them. That would be the -- what a controls-based approach might look like.

In summary, the agency has a lot of tough choices ahead, a lot of complex choices, but what I urge you to do at this moment is to start by thinking through your policy goals, and don't lose sight of them.

Notifying people of breaches is not a good policy goal. It lacks vision. Imagine if the headlines ten

years from now are largely the same as those today. We're in a kind of spin cycle of learning about security inches -- incidents and receiving notices of them that, I think, would be a failure of vision. Instead, I urge you to consider a ten-year vision for where we want the security of Californians to be. The number of security breaches that occur will obviously be part of that vision, but I would hope that the agency's efforts could palpably promote trust in digital systems, reduce the number of incidents, reduce injury from those incidents, and require collectors of data to internalize the risks they create from collecting and using data. Thank you.

MS. URBAN: Thank you very much, Professor Hoffnagle, for that characteristically pellucid explanation and presentation to us.

I am now pleased to introduce our next speaker, Professor Andrews Selbst, who will be presenting on automated decision-making, the goals of explainability and transparency. Andrew Selbst is an assistant professor of law at the University of California-Los Angeles School of Law. Professor Selbst's research examines the relationship between law, technology, and society. Drawing on resources from computer science, critical theory, sociology, and science technology and policy -- excuse me, society, Professor Selbst seeks to

understand how the creation, use, and proliferation of different technologies can interact with existing legal regimes and how legal actors can most usefully anticipate or respond to the social effects of new technology.

In recent work, Professor Selbst has focused his research on the effects of machine learning and artificial intelligence on varied legal regimes, including discrimination, policing, credit regulation, data protection, and tort law.

Professor Selbst received his J.D. cum laude at the University of Michigan, an M engineering degree, a master's of engineering degree in electrical engineering and computer science from MIT, and SV degrees in physics and electrical science and engineering from MIT. Before law school, Professor Selbst designed integrated circuits.

Welcome. We are delighted to have you here, Professor, and I will turn it over to you.

MR. SELBST: All right. Thank you so much. Let me share my slides here. Can everyone see that?

MS. URBAN: Absolutely. That looks great.

MR. SELBST: Okay, great. Thanks so much to the CEA for inviting me to give this talk today. I'm going to talk about automated decision-making and the goals of explanation and transparency in the policy responses to

automated decision-making. Often when you hear about explanation and transparency, particularly with respect to algorithmic fairness and justice, you think about things like accountability and the rule of law, right, accountability meaning people should be held -- to account for certain kinds of decisions. The rule of law, thinking of this big idea that decisions should somehow be justified or explainable. Both of those concepts are pretty vague, and explanations and transparency have a lot of different overlapping meanings in this context, and so I want to break down those different meanings.

So today I want to talk about who explanations are for; what the explanations are for; what they're trying to accomplish, which will depend on who they're for; what the different kinds of explanations and transparency are in the algorithmic context; and some specific issues that explanations can address.

So let's start with who explanations are for. There are roughly four categories of people who need explanations and transparency into algorithmic design. The first is developers. This is just basic documentation that's common to any engineering discipline. Developers require explanations, documentation, in order to -- as part of their development process. So in order to do debugging, they

need to understand, you know, what the program is accomplishing. And then there's sort of internal organizational tasks. Maybe someone gets hired and needs to pick up a project from someone that left, or teams need to coordinate together.

Again, in most engineering disciplines, this is fairly standard. I get the sense that in computer science, this is still actually developing in the algorithmic context. Particularly when it comes to discussions of the field of interpretability, most of that work has been geared towards developers themselves. I think this is a relevant separate category, and I want to get into that more when we talk about the types of explanations, but not -- but this is all focused on developers, whereas I think regulators should be focused more on the effects that sort of are outward looking rather than inward looking.

So the other three sets of people that explanations are for, I've divided into consumers and regulators, where consumers are both users of algorithmic systems and affected non-users of algorithmic systems. I separate those because consumer in the context of the CPRA refers to any natural person, right? So we're all consumers, but users of these systems have very different needs than non-users at different times. And, of course, regulators

are going to need transparency and explanations of what's going on in order to do their jobs.

So what do the explanations accomplish? Well, again, this is going to be dependent on who the particular explanations are aimed at. So for developers, it's -- again, it's internal. It's documentation for debugging, for coordination, for transitions during turnover, or even without turnover, right? Maybe somebody developed something and comes back to it three years later. This is just basic code documentation, right, that every programmer learns day one.

There's an entire field of interpretability, and this is where the understanding of the developer is sort of internal looking explanation becomes important. Interpretability is a design sort of technique or ex-post explanation technique developed by computer scientists primarily for the purpose of understanding and debugging their own algorithms. So it is -- it's almost a trope at this point to say that machine learning algorithms can be so complex, so hard to understand that even the programmers that come up with them can't understand the models. And especially for things like the neural nets, that's definitely true, but you can have the whole point of machine learning in some sense as differentiated from something that can be hand coded is that it comes up with

models so complex that it's very difficult even for the developers to hold them in their mind.

And so developers came up with different techniques, programming them in a certain way that uses a reduced number of variables or sort of ex-post interpretability mechanisms that allow them to get a better handle on even what's going on in the algorithm. A few years ago, interpretability and explainable AI were very commonly discussed as possible sort of regulatory -- avenues of regulatory pursuit to try to get at this sort of rule of law accountability idea, but it's kind of a not a great match in a way that I'll explain though in a minute, because -- in large part because interpretability is really inward looking. It's developed for developers, rather than for regulatory effects, so functionally they want validation and more debugging from interpretability.

For consumers, right, the kinds of explanations they want for users are making sure they understand what they're using and how it works. All right. So someone buys an algorithmic system and wants to integrate it into their employment process, their loan process, their Medicaid allocation, right, they need to understand what it is they're doing. This can be from a consumer protection standpoint, right. Did they even get what they were buying, does it work. All right. It's a

surprising amount of algorithmic systems on the market that just don't do what they say they do, and so there are consumer protection issues here.

There's also liability issues, right. They want to make sure that in an ideal case, right, they're not discriminating or, in a government case, that they can explain what they're doing to the people who they're making decisions about, again, with Medicaid allocations or other kinds of benefits, unemployment benefits. Are they going to be subject to challenge as government actors or private actors? So this is the kind of interests that users have in transparency in understanding their system.

Affected non-user consumers are probably the most commonly discussed people when it comes to explainability concerns and when we talk about algorithmic fairness and justice, right. So there are a couple different kinds of concerns that non-user consumers can be -- that can be alleviated by explanation. One is a question of procedural justice or the intrinsic value or explanation, right. It's a dignitarian concern. There's something just about being subject to decisions without ever being told what the basis of those decisions are that sort of strips us of dignity. This is the reason Kafka's The Trial is such a -- is a horror piece, right? It's the

horror of faceless bureaucracy that has no explanation.

Now, this is very different from the idea of contestability, but it's still contained within the idea of due process, right? Part of due process is a dignitarian concern, a simple respect for humanity interacting with their government, right? They need some sort of explanation. The same is true on the private side. It's just a dignity question.

Separately, there is the question of did this decision -- was this made correctly. Was this made in a justifiable way, right? So this is the concern of some -- usually what is referred to as due process, contestability falls in this category, and contestability is something that my co-panelists Professor Kaminski and Chairperson Urban have written about, can probably speak about more.

And the last point, right, is the possibility of enabling future success. So a lot of the times, things like adverse action notices for credit denials that are required under the Equal Credit Opportunities Act are there to say that you were denied a loan because you did X and Y or didn't do X and Y. You didn't make enough money; you changed your job too recently. The idea of those could be an intrinsic value question, but often it's seen as a way to allow consumers to adapt their

behavior to the things -- the rules that are governing their lives, right, to enable future success so that in the future they can get a loan to enable consumer choice and action. And so these are all very different reasons that explanations to consumers can be useful.

Finally, for regulators, anything needed for compliance and oversight. This is to make sure that the algorithmic systems are functional, right, to test for any sort of social and legal impacts based on whatever rules are created. So if, for example, there is a rule that creators or users of algorithmic systems have to explain something to consumers in the transparency regimes that I just discussed, then the regulators might come in to say, hey, are you setting up this rule to give the right kind of explanation. It turns out that you can have different kinds of explanations for the same phenomenon and you can justify any number of them. So you can be denied credit, again, for either, you know, too frequent job changes, or you don't make enough money, and there might be good reasons to do one or the other. Maybe if you make only a little bit more money you would have tripped the threshold to get credit, and so the justification for telling you to make -- telling the consumer to make a little bit more money, that's the thing that takes the smallest change in order for them to

get credit in the first time. That's a way -- or the next time, rather.

That's a way to explain something that is justifiable, but another one might be what makes up the bulk of the decision. Even if it's harder to change, right, if it's about respect for the decision and for the rationale, then maybe you say, hey, you change jobs too frequently, you look like a bad credit risk as a result, but that's not a way you can -- that a consumer can really respond except by staying in the same job for a long time. So what the justification is for a future explanation is something that itself needs to be justified, and that won't be explained to a consumer, but perhaps a regulator would be interested. Those kinds of behind-the-scenes questions about how decisions get made, including decisions about how to explain things, would be for regulators.

And finally, future policy learning. So I'm going to talk more about impact assessments. But the idea is algorithmic systems get tested, come with certain failure modes that -- or certain failure modes will be discovered, and we don't necessarily all know what those are and they won't be made public, but in order to understand how to regulate in the future, regulators need access to this kind of testing, this kind of

understanding, about what's going on internally in these systems, the decisions that were made, the decisions that weren't made, and why in order to understand better how -- what problems are likely to come up in the future and make regulations that are just not, you know, nail/hammer situations, where it's just, you know, do the same thing everywhere, but actually smart, tailored regulations based on realistic failure modes.

So those are the kinds of -- the goals of transparency for different actors.

Now, I want to talk about a bunch of different kinds of explanations that -- and transparency that can be enacted, some of which go to different kinds of actors, right? So in general, I tend to think of these in terms of two different sets of explanations and transparency. The first is the focus on existing models. Again, maybe four years ago this was really big in the conversation of algorithmic fairness and justice. We were talking about model explanations, right. There's an existing predictive model, how do you explain how it works; outcome explanations -- you were denied credit, why were you denied credit -- and interactive explanations where a consumer can go and sort of play with a model and get an intuitive sense about how things are moving within the model.

The issue with these explanations is fundamentally they take the existing model as a given and therefore put the onus for change, for challenge, on the consumer themselves, which is -- who are often powerless or relatively powerless. And so more there's been a move towards a focus on model development, right. Model development includes documentation, impact assessments, audits; it's really more of a focus on the people who are creating, implementing, and using these algorithms. Both have their place, but I believe today that the regulatory focus should be more on the people who are creating and using the algorithms, because again, they have more power to change the reality on the ground than individual consumers do.

So let's go through it. Explanations of existing models, right. So the first big one is outcome explanations. They're targeted at affected consumers. They can enable future actions or appeals. Again, you were denied credit; here's why, right. Hopefully they can tell you enough to know should you appeal this decision. Was it made illegally. If you get an explanation of the law or if you can take your explanation of the decision to a lawyer who knows the law, right, you can get an answer on whether it was illegally made and you have a case for appeal.

They can be dignity enhancing. Remember, these are the ones -- these are the explanations that go to people who have decisions made about them, and so they are necessary just for basic dignity. But they can be often -- they can be underspecified in a way I discussed before, and they can be easily manipulated, right. If we have a weaker explanation-focused regime that does not dictate exactly how explanations are given or gives a lot of leeway, you can get somewhat meaningless explanations. So in the case of the Fair Credit Reporting Act, the adverse action notice regime, there's a lot of explanations, reason codes, given that don't help, right. Some of them say, hey, there's no existing credit file, which doesn't say much about the credit determination at all, but it's a very useful actionable item. Some say, you know, length of time at job, which doesn't even say whether it's too long or too short, and it doesn't tell you how long -- it's not actionable and doesn't really say much. The reason codes given in that regulation, regulation B, are meant to be a sample, but are often used wholesale. So we have to really think about, if we focus on outcome explanations, what specific goals we are meaning to achieve and lay those out in a very concrete way.

Model explanations take an existing model and try to

simplify it in a way that's easier to describe, right. So either you take a localized outcome, right, so consumers who kind of look like you, what is different. So if you vary your income a little bit and it's -- the model is less sensitive to that than how long you've been at your job, then you understand more about the model, right. It's more sensitive to one variable versus another.

You might be able to have a picture of how the model works in a localized way, but you can also have something where it simplifies sort of a deep neural nets into a decision tree, right. That's clearly simpler, but those are often still too complicated to understand in any meaningful way. A person can't hold them all in their mind. So model explanations, right, because a decision tree with a thousand branches is not something that a person can understand enough to act on. They can trace it on a chart and literally follow the answers, but you can't explain it in a way that's helpful to consumers. And so model explanations are often in the category more of interpretability that is geared towards developers in order to help them sort of debug.

And the last is interactive explanations. You see this on things like creditkarma.com where you have a drop-down menu. If you increase your employment by

X -- or your salary by X amount, if you pay down certain amounts of debts, what will happen to your credit score, right, things like that. Interactive models where people can get more of an intuitive feel. This is helpful. You don't actually need explanations to be literal, you know, language-based explanations, and intuitive feel is helpful, but it can be very misleading, especially when models are non-monotonic or not even continuous, right. You can end up with consumer's sort of playing around with an explanation here and then it drops off a cliff in an area they didn't have access to. They can get a very misleading picture of how a model works. And so all of these explanations of existing models, to the extent they should be useful to a regulatory regime, put all the onus on the consumer in a way that can be somewhat empowering, but also quite misleading at times, which is why I say that the focus is better on -- is better put on explanation of the model development process.

What I will say is current law, where it requires explanations, is more focused on existing models, right. So I've mentioned adverse action notices a couple of times. There's also the GPR, which again, Professor Kaminski will speak about a lot more. But here, I'll say that article 22 requires safeguards for automated processing, including human intervention and

contestation. The only way you get contestation is an explanation that enables you to know when to contest something, right. And similarly, article 13 to 15 of the GPR, each have a subsection that requires meaningful information about the logic involved in automated processing.

And again, what does meaningful mean? Well, this has not been litigated. It's not -- it's not obvious what meaningful should mean. I believe it's functional. I believe meaningful should imply the ability to enforce your data subject's rights under data protection law or human rights law. So it means, again, contestability on the basis of discrimination, on the basis of illegal processing of data, things like that.

To the extent we should have consumer-focused explanations, I think a functional -- functional test that asks the degree to which the explanation helps the consumer enforce other existing rights is probably the best approach.

Finally, we get to explanations of model development. And here I have, again, three different kinds. One is documentation, and then impact assessments and audits. Again, documentation is just basic, right? Basic standard practice in engineering. It allows coordinating; it allows handoff. One thing documentation

should do is describe the limitations of the product, the failure modes, the testing that's been done.

One of the big models that's floating around out there in the -- the computer science world is drawn from electrical engineering. It's called data sheets. And here I -- this -- I'm partial to it because of my history as an integrated circuit designer. Whenever you buy a chip, it says, it works under certain conditions. If it gets too hot, the response curve falls off. If it -- you know, if it -- if the signal is too fast, you're not gonna get a good response, right? This is not to say that the chip doesn't work. It just says, here are the conditions under which it works. That's the idea for data sheets, right?

Similarly for this. If you train a model, if you use -- if you give a certain data set, you should say, hey, here are the conditions we've tested, here are the conditions under which it works. We've used this -- we've trained it with a data set that this demographic shape. If you try to deploy it on a different one, it's not going to work. But it's still saying the product works, but it specifies the sort of sphere in which the product works.

There are moves within the sort of algorithmic space to make this much more common, to -- to make benchmarking

decisions and -- and -- and documentation just standard practice, and those absolutely should be, which, from a regulatory side, actually works very well with things like impact -- impact assessments and audits.

So impact assessments, the idea of them is to document important decisions before they're made, right. To predict social impacts. So I'm -- early in the process, I'm thinking about developing algorithmic -- the system. I want to figure out, what are its limits, what kinds of future social impacts will it have, do real, rigorous research, and then figure out how to mitigate these before even going -- going forward with the development.

The earlier in a process that you can count social values and embed social values, the better for the ultimate social harms, or mitigating the social harms of a product. This arguably does exist already in impact assessment regime in the GDPR, right? In Europe they have Article 35; it requires data protection impact assessments whenever you have high risk processing. And that could be considered an algorithmic impact assessment. Similarly, right now, there's pending federal legislation, the Algorithmic Accountability Act of 2022. It talks about impact assessments. Canada has implemented a version of impact assessments. And so this

is coming globally, right? This is becoming a response that is common -- or commonly proposed and, I think, incredibly important. The big thing about this is it's not necessarily about explaining existing models but about explaining the decisions that went into them, right, what data sets were and were not used to train the models and why, what were the goals, the optimization criterion, and why were those chosen?

One of the major benefits of it, again, is an agency like the CPPA can receive a whole lot of these impact assessments, and then over a year or so, do some analysis and see what works, what doesn't, what causes problems, and learn from them.

Similar to impact assessments are audits. There's a lot of disagreements about whether impact assessments or audits, like, where the two sort of differentiate from each other. I think a lot of people think of audits as similar to impact assessments. So audits can be internal audits. Often engineering firms will just say, hey, before this goes out, let's do a check. Let's make sure all these -- these harms -- potential harms are taken into account. They can be external, independent audits, right? I think in the legal space, we think of external audits as much more rigorous and important because it's not pure self-regulation. They can be mechanical, an

audit of the system, the -- or they can be regulatory, which is more an audit of the business practices. And so there's a -- audits are sort of an all-encompassing idea. I think the biggest difference in my mind between impact assessments and audits is impact assessment are bottom-up. They ask you to document your decisions, to say, hey, why did you do this? Audits are usually, as I understand them, top down.

They're saying, hey, these are the -- what we know are best practices. Did you do these things? They often turn into a checklist. But in order -- and a checklist is not ideal. But in order to even get there, we need to have a very good sense of what it is we're saying should be done. And I don't think we're necessarily there with algorithmic systems yet. We don't all agree on the different sort of standards for -- for harm that's tolerable or -- or what counts as discrimination. And so until -- I think the impact assessments are a way to get us to learn enough to maybe have a more rigorous audit regime that can focus on concrete things we know are harms and how to avoid them. But we do need some combination of the two. And again, because they're both focused on business practices, they can be much more useful than ex-post explanations.

Finally, in my last two minutes -- I don't know how

I'm doing on time, but I'll try to keep this last bit quick.

MS. URBAN: You are fine. Thank you.

MR. SELBST: Okay, great.

So where -- I want to focus on specific issues that documentation can address, right? And explanation -- and the truth is, right, explanation and transparency without -- for its own sake isn't particularly helpful. It needs to be tied into a sort of reform or accountability goal. And so we should keep in mind what the things are that we're trying to achieve with any transparency, any explanation regime that we are trying to implement.

So I probably don't have time to talk about all of these, but I want to talk first about the question of whether something works. Again, there -- there's a surprising amount of AI that's being put out under this sort of -- this hype umbrella of AI, that simply just doesn't do what it says on the tin. And transparency as to what people are thinking they're trying to do, and allowing people to test whether it works as it says, is just sort of basic, right? We've had that for every product that's ever been on the market since snake oil. And so that is a basic reason for transparency.

Another issue is divisions of responsibility. And

so to draw on a -- an analogy that I think is familiar to all of us, the question of cars turned out to be a very difficult one for tort law to address from a responsibility standpoint, right? So here you have automakers that are making cars, and users, drivers, that are driving them. And they'll get in accidents, right? And drivers can, you know, be sued in negligence when there is an accident. And it is either not their fault, because someone else caused the accident, or it is the driver's fault, because they were negligent.

And for a long time, manufacturers of auto vehicles tried to say, well, look, it's not our fault the car got in an accident, even though the passenger died, right? And the passenger might not -- maybe there weren't airbags, right? The passenger need not have been as injured as they were, or the driver. As documented by Professor Bryan Choi at Ohio State when he's talking about software, eventually tort law came up with this -- or judges came up with this crashworthiness idea, which says, look, you, the auto manufacturers can't be totally off the hook, right? It's not an unknown thing that car crashes will happen. You don't know whether any individual car will crash or when, but you know for certain that car crashes will happen. And so you need to be responsible for that second level of harm, right? If

someone gets hurt worse in a car crash than they otherwise would have been, that should lead to responsibility for the manufacturer, as well as responsibility -- any blame that goes to the user.

So there's a question of divisions of responsibility. And in order to -- in order to know how responsibility is divided, they need to understand the kinds of testing, the kinds of expectations in the design of the product that were built in, right? Whose responsibility is what? What should the user be responsible for? Those questions can't be answered unless you have some visibility into questions of design.

The same is definitely true in algorithmic design, and artificial intelligence, right. If you imagine an employer that creates or hires a software developer to create an algorithm that discriminates, right, the degree to which they're in communication about how to solve the problem will tell us a lot about how to allocate responsibility between the two. The framework where the developer, like, cabined their problem, right, did they take the particular demographics of the training set versus the deployment set into account? Maybe that's a developer problem; maybe that's a -- a -- an employer problem. But we don't know unless we have transparency into those design decisions. And so it's really

important for divisions of responsibility between users, either affected users, right, if they have agency, users who are consumers, and developers.

Finally, in a paper called "Fairness and abstraction in socio-technical systems," I with several co-authors discussed several abstraction traps, which are common mistakes that developers are making when trying to create fair machine learning algorithms. I want to talk specifically about two of them. The others can be found in the paper. The five abstraction -- or the two I want to talk about are the portability trap and the formalism trap, because I think both are very, very common.

The portability trap is, again, this idea that you can train for one context and deploy to another. Maybe you train an algorithm in a medical context and deploy it in a -- in a prison, right? Or you train an algorithm in Tennessee and deploy it in New York City, right? The differences between those, right -- computer science is very focused on creating modules that are abstract and able to be redeployed elsewhere. That is -- it's an aesthetic sensibility. It's something you start with day one in computer science, learning about abstraction. But the problem with it is that a lot of algorithmic systems functionally are -- function based on database taken into account, which has a context. And when you strip it from

the context, you cannot guarantee any sort of fairness. And so again, in order to understand whether you're making this error, you need to have transparency into the sort of justifications for porting it from one to -- one context to another, or the kinds of mitigation you've got.

Formalism trap, here we talk about how a lot of fair ML systems are trying to describe discrimination as a sort of mathematical formula, which will inevitably cut out a lot of the nuance of what lawyers and philosophers and sociologists mean by discrimination. Sometimes it can work, but in order to make sure it works, you need to be very, very specific about your rationales for modelling -- for modelling discrimination in this particular way, in this particular context. And again, the decisions that go into building these systems need to be able to be evaluated.

So with all these abstraction traps, in fact, all these issues, right, they're all about algorithm design and the decisions that went into them. So what I would say is, as you think about the regulation, the focus should definitely be on questioning how algorithm systems are designed, how harms are evaluated, how they're mitigated, and eventually, whether they work if they are deployed. All right. Thank you very much.

MS. URBAN: Thank you very much, Professor Selbst. Again, that was very helpful and clear, and we really appreciate you doing this for us. So thank you.

I'm very pleased now to introduce our final speaker, Professor Margot Kaminski. Professor Kaminski is an associate professor at the University of Colorado Law School and the director of the privacy initiative at Silicon Flatirons at the University of Colorado. She specializes in the law of new technologies, focusing on information governance, privacy, and freedom of expression. Recently, her work has examined autonomous systems, including artificial intelligence, robots, and what we commonly know as drones. In 2018, Professor Kaminski conducted research on comparative data privacy law as a recipient of the Fulbright Schuman Innovation grant. Her academic work has been published or is forthcoming in Columbia law review, the UCLA law review, Minnesota law review, Boston University law review, and Southern California law review, among others.

Prior to joining Colorado Law, Professor Kaminski was an assistant professor at the Ohio State University Moritz College of Law and served for three years as the executive director of the information society project at Yale Law School, where she remains an affiliated fellow. She is a co-founder of the Media Freedom and Information

Access Clinic at Yale Law School. She served as a law clerk to the honorable Andrew Jay Kleinfeld in the Ninth Circuit Court of Appeals, and she holds a JD from Yale Law School and a BA from Harvard University.

Professor Kaminski, welcome, and the floor is yours.

PROF. KAMINSKI: Thank you so much. So give me one second here to share my screen. So I'm gonna be presenting today on automated decision-making, and some of what I'm going to say here is going to overlap with what Professor Selbst introduced to us. But my perspective is a distinctly comparative one. And what I'm hoping to do with this is to make clear to you the influence of the general data protection regulation from the EU on some of the language that is in the CPRA, and now CCPA, and also to talk a little bit about the model that it creates and what problems that model contains and what benefits.

So my presentation will consist of three parts. First, I'm going to discuss the actual law and provide the legal background for this comparative perspective, for those of you who might be less well versed in the GDPR. Then I'm going to talk about comparisons between the GDPR's model and several other models that are out there, primarily in the United States. And finally, I'm gonna provide some normative takeaways. So with these

takeaways, I'll be drawing on the legal background I've provided in parts 1 and 2, and also on some other comparative work that I've been doing. And some of this will definitely resonate, again, with what we've heard from Professor Selbst.

So in a nutshell, again, for those of you who might not be so familiar with it, the General Data Protection Regulation, which is the large-scale data privacy or data protection regulation in the European Union, consists in a nutshell -- this is probably the shortest presentation I'll ever give on the GDPR as a whole -- of two parts. There's an individual rights section, many of which are going to be familiar to you as reflected in the same individual rights in the CCPA. And then there's a section on data controller, which for our purposes really means company, though it also includes government entities in Europe.

Data controller obligations. This part, the second part, is largely missing from many US state privacy laws. And we'll see how, today, how some of the influence of the GDPR is making its way across the pond to data controller or company obligations.

The second thing to know about the GDPR, from a bird's eye view, is that it has a very specific governance style, which is gonna be relevant for my

discussion of impact assessment in particular. The governance style of the GDPR is that it consists of often vague tests with high -- tests with high level concepts, such as the concept of fairness or the concept of lawfulness or the concept of discrimination.

And then that text is delegated, in terms of its interpretability, to a number of different possible actors. Some of those actors are regulatory actors. So the European Data Protection Board issues guidance, which is not formal law, but certainly helps interpret what the text of the GDPR means. And some of those actors are nonregulatory actors, namely the companies that are doing the implementing of the GDPR on the ground.

So when you look at the text that I'm gonna talk about today, or you look at the text that Professor Selbst mentioned already in the access and notice portions of the GDPR, when you're asking, what does something mean, if there's not an answer from the European Data Protection Board or an answer from a court, the answer is, the company implementing the law is going to be deciding what that means.

So this is a largely collaborative, and deliberately so, method of regulation, which might be surprising to some US persons who look at the GDPR and think that it's a top-down regulatory control version of privacy. GDPR

also contains a number of very explicit collaborative compliance tools, such as certification and various sort of collaborative codes of conduct that can be created. But that collaborative governance is done in a very specific European context, which, as Chair Urban knows well, exists against the backdrop not only of huge regulatory fines but against a human right -- a right not only to privacy, but to data protection, which is interpreted by not really one but two human rights courts.

So where you have a vague term, and it's being interpreted by a company and application, the company's still bounded by the fact that a human rights court may step in at some point to specifically interpret the term and help regulators enforce it. GDPR, unlike a lot of the state laws that we've been seeing, has both private and public enforcement, although class actions are not an instrument that really exists in Europe. There's an attempt at sort of a joint action in the GDPR.

And additionally, as I said, it has these regulators who have been in place since long before the GDPR, who both provide guidance and often do the enforcing. Plus, the GDPR must be understood against the backdrop that many EU member states have had in places laws that have existed since the 1970s, and the regulatory

infrastructure for enforcing them has existed since similar times.

So turning to automated decision-making in the GDPR, the GDPR on the whole regulates the processing of personal data, wholly or partly by automated means, in addition to the processing of personal data in a filing system. That means that any time the data is processed -- personal data is processed by automated means, it's gonna be subjected to the whole of the GDPR. Additionally, and perhaps confusingly, for those of us who are sort of looking at it more myopically from this side of the pond, the GDPR also contains specific provisions on automated decision-making with significant effects.

So in my presentation on this first part of the actual legal bases for regulating ADM in the GDPR, I'm gonna start with these ADM specific provisions and then briefly point to several generally applicable provisions that are going to be relevant as well. Before I get into this, I just want to again really briefly talk about what the GDPR is aiming to do.

So I've identified this in a piece that's called, "Binary Governance", in which I talk about how this two-pronged system of individual rights combined with compliance infrastructure or governance aims at three

different goals with respect to automated decision-making. And again, this echoes with Professor Selbst's presentation.

The first goal is really an instrumental one. So the instrumental goals largely look at -- largely sound in the idea that we are trying to fix automated decision-making. We're trying to prevent errors; we're trying to get rid of discrimination; we're trying to watch out for places where the ADM might crash or produce incredibly unexpected results.

A second goal of regulating automated decision-making is a dignitary goal. And again, this is something that often gets characterized as being more European in nature, the idea being, we don't want to, you know, take away your name and give you a number or turn you into a data double and objectify you. However, there are certainly dignitary conceptions that echo in regulations in the United States. Professor Selbst mentioned, for example, the FCRA.

Third, and pretty significantly, the GDPR is concerned with lawfulness. It's concerned with accountability. And so when it's talking about data protection, it's not necessarily talking about the kind of privacy many of us think about when we think about being left alone. It's talking about power and

accountability and disparities and access to data and access to -- to decisions, and power over individuals through files that are held on them.

So justification really goes to this idea that the entity that has the power, the entity that's using the automated decision-making based on personal data, needs to both provide individualized explanations that justify its use and show that it's making socially normatively okay decisions. And additionally, the system as a whole needs to be justified as legitimate. We'll talk about more about that in a minute.

So going to the actual text of the law, the GDPR contains, what should be very familiar, a series of notice rights and access rights for individuals. The individual notice rights contain within them a right to notice about an automated decision with significant effects. These exist in Articles 13 and in Articles 14, one of which deals with gathering data directly from individuals and the other of which deals with gathering personal data from third parties.

The company or controller has an affirmative obligation to disclose not only the existence of automated decision-making, the fact that it exists, but also meaningful information about the logic involved. And there's your language from the CPRA. And

additionally, the significance and envisioned consequences of the processing for the data subject. Basically, why does it matter to you? Right, what's the consequence of this going to be.

As to timing, the timing when you're collecting data from an individual directly is supposed to be at the time that that data is collected. The timing when you are gathering data from a third party, under Article 14, is that it has to be within a reasonable amount of time, which, in the text, is supposed to be no longer than a month.

The second set of information rights around automated decision-making come from the GDPR's governing of individual access. It's the exact same language -- which might have interesting implications, by the way, for how one interprets the phrase "meaningful information" about the logic involved. But in Article 15, a person -- an individual requesting access to information from a company must be provided with meaningful information about the logic involved and the significance and envisaged consequences of the processing.

In terms of timing, the individual can ask for information easily and at reasonable intervals, and yes, may be charged for it, but only under certain

circumstances. This brings me to Article 22. So Article 22 is probably one of the most discussed and least understood portions of the GDPR. It established what is, in effect, automated decision-making due process. That is, it gives individuals who are subject to automated decision-making with significant effects the ability to contest such decisions. The start of Article 22, however, does not look like due process. It looks like a ban. So it states that the data subject -- that's the individual -- shall have the right not to be subject to a decision based solely on automated processing which produces legal effects concerning him or her, or similarly significantly affects him or her.

This, interestingly, given all the focus on automated decision-making and accountability recently, is not at all a new right or set of rights. It is based on the language in the previous European Data Protection Directive, though it has some changes that, in my view, make protections of Article 22 broader, deeper, and stronger. That is, it applies to more types of processing, it creates more restrictions, and it's backed by more significant enforcement capabilities than its predecessor.

For example, the guidance from the European Data Protection Board suggests that the terms based solely on

automated processing does not leave out situations where a company adds a human in the loop solely to try to escape Article 22. So if you have automated decision-making and you put a human in it to try to get out of this -- these restrictions in -- or governance in this particular provision, you're not gonna be able to do that, at least under the guidance from the European Data Protection Board.

Second, and significantly, the guidance suggests that similarly significant eff -- effects can actually be quite broad. So one way to read this -- one previous way to read this provision is to say that legal effects are fairly narrow and include only things such as housing decisions or employment decisions. But the guidance from the European Data Protection Board suggests that things like manipulative advertising, when it's particularly egregious, could be covered by this as well.

Finally, the guidance establishes -- and this is actually important just in terms of the requirements and how many of them take effect on how many companies -- that this is a ban and not opt-in. So a number of the member states that implemented the previous version of Article 22 from the European Data Protection Directive read it, or really implemented it, to be a right that individuals had to opt into, which meant that the

restrictions it places on companies only really took effect if a person opted into the right. The guidance here makes -- makes clear that that is not the case.

So we have this ban, right? We have this ban that says, don't use solely automated decision-making, which really can mean decision-making with people that's automated, that creates significant effects on individuals; don't use it, except the exceptions. And the exceptions actually end up being much of the rule.

So if an individual gives explicit consent to automated decision-making -- and what explicit consent is, is rather debatable, but it's considered to be even more heightened than the GDPR's already strong protections for consent -- if they consent, or if it's necessary for a contract, or if as has already happened, a member state creates a law that authorizes particular forms of automated decision-making, then a company can use -- or government agency can use automated decision-making with significant effects on an individual.

That's not the end of the story, though. So when they are using automated decision-making, they are then required to put in place what are called suitable measures or suitable safeguards. So a data controller must implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests.

So the -- the real meat of the GDPR's regulation of automated decision-making, or AI, is this question of what constitutes a suitable measure or a suitable safeguard. And remember, back in the beginning of this part of the presentation, I talked about how the GDPR really has these two prongs. One side is individual rights, and the other side is compliance, while on the face of it, Article 22 looks like it's all about individual rights. It looks like it's algorithmic due process.

But the actual content of Article 22, as interpreted through various instruments that accompany it, shows that it is significantly more than this. Within the text itself, the safeguards that are listed -- which are really -- it's an open list, not a closed list -- but these are mandatory -- include individual due process -- that is, the right to obtain human intervention in a decision, the right to express one's point of view, and the right to contest or challenge the decision. None of this is operationalized in the face of the GDPR.

What I find to be really interesting about the provisions that you are tasked with interpreting is that there isn't an equivalent in the CPRA or CCPA. Instead, there's a right to object. So one question for you in terms of trying to figure out how much alignment you want

with what companies are already having to do under international law, or under transnational law, is to try to figure out whether your version of objection is gonna map onto this algorithmic due process that's in Article 22.

The suitable safeguards, when you look at the recital, Recital 71, that accompanies Article 22, also clearly indicate that there is an individual right to explanation. Now, there's been debate over this. I think the debate in my view is rather silly. It's very clear that in the accompanying document that goes along with the text, that regulators have decided to interpret Article 22 to include an individual right to explanation.

In the guidelines, too, the European Data Protection Board points out that this right to explanation is not necessarily the same thing as the disclosure of meaningful information about the logic involved. The Article 22 right to explanation is clearly outcome-geared in the sense that it's trying to enable an individual to exercise the other rights. So an explanation must allow the individual who's affected by the decision to be able to challenge that decision. Whatever you put in the explanation needs to enable the rights.

But it's a big mistake to think that these individual rights are all that the GDPR has to say on

automated decision-making. Article 22 clearly, in its suitable safeguards or suitable measures, also aims at creating systemic compliance through risk regulation with a number of substantive goals. Recital 71 states that companies should implement technical and organizational measures to ensure that algorithms are not inaccurate, to make sure that inaccuracies on a systemic level are corrected. Recital 71, which again is the interpretive text that accompanies Article 22, not strictly speaking hard law, but certainly soft law or guidance, says that companies must prevent discriminatory effects on natural persons on the basis of a whole long list of categories, some of which are familiar to us in the US, other ones, like trade union membership, are more specifically European in nature. And both of these provisions suggest that companies are ex-ante responsible for the system and for ensuring that the system doesn't result in certain kinds of predictable failures, the harms of which sound in individual rights again.

Furthermore, the guidelines on automated decision-making say that audits are recommended as part of suitable safeguards, that third-party expert oversight is recommended as part of suitable safeguards, and finally, point to the impact assessment as being part of this systemic regulation.

So this brings me to the connection between the individual due process rights to the GDPR, and these impact assessment, the DPIAs. First, it's important to understand that DPIAs apply well beyond algorithmic decision-making in the GDPR. They apply to data-processing in general. Automated decision-making with significant effects is identified explicitly in the text of the GDPR as one type of high-risk data pro -- processing that requires a DPIA. So if you understand Article 35 as establishing that any high-risk processing -- which is a standard, not a rule -- must be subject to a DPIA, and particularly high-risk processing must receive regulatory oversight before the -- it's actually deployed, then automated decision-making is a rule within that standard that at least qualifies for the requirement that you conduct an impact assessment, and possibly sometimes qualifies for the requirement that you also consult a regulator before you release the algorithm for use.

The guidelines on automated decision-making emphasize the centrality of the DPIA, as have several scholars. Most specifically, I'm thinking of Michael Veale and Lillian Edwards. They say that the DPI, the impact assessment, is a crucial aspect of suitable safeguards, that it aims at systemically mitigating harms

on a systemic and importantly, ongoing, not just ex-ante basis, to individual rights and freedoms.

So what's interesting about this is that you have a risk regulation process, right? You have an ex-ante impact assessment conducted on a systemic level. It's gonna try to mitigate harms. And traditionally, when we think about risk assessment, or we think about risk mitigation, we think about things in safety-critical contexts. So we think about, you know, mitigating the harms, or preventing the harms, like Professor Selbst said, of car crashes, right? You put in more airbags so people get hurt less often.

The big trick here, the really difficult thing, is that the harms that are being mitigated in an impact assessment, in this context, are often not measurable. Not only that, they're often contestable. So they are harms to individual rights, not necessarily harms to something that is quantifiable, measurable, or physical. So with my colleague, John Claude Malchieri (ph.), who is a professor in Europe, at Edhec (ph.), we have looked at this and said, there's something about the DPIA specifically that connects to this -- the GDPR's due process rights. And the DPIA, as conceived of in the GDPR, when it's applied at least to automated decision-making, should feed into the kind of information that's

disclosed to individuals. They're symbiotic, these individual rights and this systemic analysis.

Finally -- I know I'm still in part one of the presentation, but the other parts are much shorter, I promise -- it's crucial to understand that these rights, the rights of notice, the rights of access, the rights articulated in Article 22 and the obligations and the DPIA, do not exist in a vacuum. DPIAs exist for other forms of processing, including systemic large-scale surveillance of public spaces. So to the extent that those data sets end up feeding into automated decision-making, they are additionally governed under the GDPR.

The GDPR proposes high level principles that are enacted throughout the regulation, that also have bigger implications for data processing, even if it doesn't fall under Article 22, and that includes fairness and transparency, purpose limitation, and data minimization -- that is, state why you want the data and how you plan on only using it for those purposes -- and accountability, which I'll come to in a moment.

There's a substantive requirement of data protection by design and by default. So if you're designing an automated system -- automated decision-making system, or really any profiling of individuals, you have to take into account these principles and design your technology

ex-ante so that it actually is built to execute those. And there's a series of under -- other individual rights, including, for your purposes, an actual explicit right to object. Now, as I pointed out, in Article 22, you have a right to contest an automated decision. But additionally, you have a right to object to processing writ broad (ph.), not fully broad. There's some restrictions on it. But this includes and goes beyond automated decision-making to other kinds of processing.

Often, that right to object involves the balancing test that is conducted by the company. But in at least one case, it includes an absolute right to object to direct marketing. So there, if you object to the use of your information for direct marketing, then the company can't decide that it's outweighed by other interests.

So part 2 of this presentation, I want to talk a little bit about some normative comparisons and comparative observations about the GDPR's mode of governance of automated decision-making, compared to what we're seeing arise in the United States. So the first thing I want to mention is that the GDPR's approach to impact assessments is not really truly exportable without understanding what else is going on in the GDPR.

So it's very much situated in this two-prong approach of rights and compliance, and in the

collaborative governance nature of the GDPR. That is, it's, on the one hand, systemic risk regulation, which we know how to do very well in the US. But on the other hand, it's systemic risk regulation that is targeted at protecting for human rights that are elaborated by the human rights court. Stylistically, it's meta regulation or collaborative regulation, the idea being that much of what happens in the DPIA is geared at trying to affect the internal infrastructure, the norms, the heuristics of a particular company. And accordingly, it really relies heavily on the regulatory infrastructure that exists in Europe and certain sort of norms that Europeans have around collaborative regulation.

By contrast, proposals to regulate ADM in the United States notably largely include -- exclude individual rights. So this is one place where I think you have the capability to really be a norm entrepreneur in this space, where the right to object to an automated decision-making would be one of the first, if not the first, examples in the United States of an individual right that is granted in the context of automated decision-making. And what I point out is that you might be a norm entrepreneur, or the California Legislature might be a norm entrepreneur, citizens of California might be a norm entrepreneur in this sense, but they're

not on the outside, internationally. The international trend has been to recognize a right to contest automated decision-makings, even if such a right has not really been proposed in the United States.

The impact assessments that have been proposed in the US have largely had a very different flavor to the ones that are proposed in the GDPR, even if you can kind of track some similar -- some similarities between them. What do I mean by this? The impact assessments in the US that have been proposed around automated decision-making are largely envisioned as more of an enterprise risk-management tool than as a collaborative governing conversation with a particular regulator and the public. That is, they're characterized as self-assessments that are largely aimed at internal risk mitigation.

There is, however, a third model for impact assessment, and I believe Professor Selbst referenced this as well, that neither the US nor the EU seems to really be following, with the exception of really one law that was proposed in Washington state. And that's MEPA (ph.), with the environmental impact statement. And this is to use impact assessments, not just as internal risk management, nor just as collaborative governance in conversation with a regulator, but as iterative policymaking and a form of public accountability, so that

the public, or at least impacted stakeholders, can see what's going on and influence policymaking further in this space.

The recently dropped Wyden, Booker, Clark Algorithmic Accountability Act tries to kind of thread a needle between these three models. So some of the impact assessment is considered to be just self-governance and internal and enterprise risk management. Some of it is actually clearly collaborative governance, in that the reports go to the FTC, and the FTC actually manages -- under this model, reports to the general public, including a partial publicly disclosed database. And then as I mentioned, this proposed Washington law, SB 5116, does actually require full public disclosure of impact assessments. They call them algorithmic accountability reports. However, this is done only in the state actor space, not in the private sector.

So this brings me to two major weaknesses for the GDPR regime. The first is public accountability, which in my view, at least in our country, is absolutely necessary for some sort of collaborative governance with the private sector, and the second is stakeholder participation. That is to say, voices in and voices out, right? The ability of individuals to influence the -- the creation of policies around impact assessments when

they're actually being harmed by the algorithm.

So the GDPR does say that companies have to consult data subjects. But that shall is largely modified by the requirement that they only need do so where appropriate, and it's limited by concerns around trade secrecy, et cetera. And the guidelines on this say that consultation could be as simple as basically a Qualtrics survey. And we all know how incredibly informative Qualtrics surveys can be. Additionally, DPIAs are recommended to be made public in the GDPR, but certainly not required.

And so this is this place where you have a delegation by a regulator to a company to say, you know, analyze your system to make sure it's not discriminatory. Anybody who's impacted by that discrimination doesn't have to be in the room when you try to figure out what discriminatory actually means. And then we have no real oversight except for spot inspection or in some cases regulatory inspection -- regulatory pre-approval of whether you've actually effectively mitigated these harms.

That is, voices in and out are crucial for both defining the nature of the contested harms that are to be mitigated, making sure that we have a -- a not just company-defined definition of fairness, discrimination, bias, error, et cetera, and also making sure that the

system as a whole is not just captured by an individual private company conducting what's effectively self-regulation.

So by contrast, there's actually some interesting things to be learned, again, from these proposed US laws, in other jurisdictions or federally. And they take stakeholder participation much more seriously, even if, as I mentioned, they're not really leaning on public disclosure of impact assessments. There's one system, one proposed law in Washington, the same one, that suggests that a regulator must consult with affected communities during the rulemaking process. And in the Algorithmic Accountability Act, proposed Algorithmic Accountability Act, a company is not required to consult with affected communities and representatives but must chart its consultation with affected communities and representatives and explain why it hasn't taken those suggestions into effect for each individual impact assessment.

And again, this is necessary not just for oversight over the actual algorithm, the technology, but also for oversight over the process by which the company is effectively self-regulating or mitigating the risks.

So to my takeaways, and to close, the GDPR, and for that matter, the draft EUIA Act, which I haven't had time

to get into today, first doesn't really raise a lot of big definitional concerns about what we call automated decision-making. And that's because there are other parts of Article 22 that serve as gatekeeping functions. So the regulations of Article 22 are triggered less by are you a solely automated decision-making system, and more by the fact that such decisions have significant effects. Or the DPIA requirement is triggered really by whether there are high risks from processing. Thus they don't really need to sort out what counts as an automated decision.

A second takeaway -- this resonates with Professor Selbst's presentation as well -- is that the EUAI Act differs from the GDPR in that it largely focuses on the producers of the technology and not as much on the users. And the developers and users of automated decision-making should really share responsibilities for those harms. Again, Professor Selbst gave the same analogy I was going to use. It's like thinking about the driver of the car and the car manufacturer. On the one hand, the driver of the car knows the context in which the car's being deployed and should have responsibility for deploying it in harmful manners. On the other hand, the car manufacturer's not off the hook, because if you design tires that can't work in winter weather and you don't

provide other options, then the car is gonna crash.

Third, as a takeaway, I wanted to speak just very briefly to a human in the loop. Placing a human in the loop, which is how some people read Article 22, but not how I read it, is the least sophisticated and probably the most problematic mode for governing automated decision-making. That is because the humans that are placed in the loop are rarely empowered, and a hybrid human-technical system creates human factors -- engineering tell us this -- plenty of additional challenges of its own, like the handoff problem. How do you alert people, how do you train people, how do you keep them engaged? And so this dominant model that's emerging international is this combination of systemic oversight coupled with robust individual rights, which, again, I think you have the opportunity to make real here in the United States.

Third -- fourth, while risk regulation is the dominant model, there is significant challenges with using risk regulation alone, so impact assessments alone, to regulate automated decision-making. And this comes to the nature of the harms primarily, though there are other issues as well. The harms in automated decision-making are not quantitative or quantifiable physical risks, but are contested concepts, such as discrimination and

fairness. And this implicates not just, you know, the possibility of trying to define these harms in advance through regulation, but also how we design regulation so that other actors, not just companies, have input into these contested concepts. In the European Union, this involves a human rights court. And here we have to think really critically about how to involve impacted stakeholders and what regulations might look like.

Fifth, an impact assessment can be a very different tool in different regimes toward very different goals, *depending* on when you're looking at it as an instrument of meta governance to get companies to change their heuristics, versus simple enterprise risk-management, versus the NEPA model of public accountability and policy and duration. And an impact assessment can take a lot of different shapes with respect to time. It can be ex-ante, it can be ongoing, it can be iterative, it can include or not include post-market measures.

And finally, finally, most crucial, voices in and voices out are essential to effective governance. Transparency matters not just because the algorithm itself is part of this trope of the black box algorithm, but because transparency makes the difference between an impact assessment being a self-assessment versus being actual governance. And these AIAs and DPIAs can be

linked to individual disclosure rights to provide some of that transparency. And finally, impacted stakeholders must be involved.

Thank you very much for your time.

MS. URBAN: Thank you so much, Professor Kaminski, again, for that really interesting, informative, and clear presentation. It is much appreciated.

I would like to -- Professor Kaminski was actually our last speaker for the day. So I would like to just take a moment to thank all of the speakers over day 1 and day 2 again for developing the deep expertise that they have and for being willing to take the time to share it with all of us. As a reminder, of course, the guest speakers' views should not be taken as the views of the agency or the Board. They are the presenter's views only. But I hope that they were interesting and informative for everyone listening.

We will now proceed with public comments on any of the presentations for today. I will go ahead and say a little bit again about how the mechanics of public comment work, just so everybody is comfortable with that, and then we will go into it. If you wish to speak on an item, please use the -- excuse me, if you wish to speak, please use the "raise your hand" function, which is in the reaction feature on the bottom of your Zoom screen.

The moderator -- our moderator, Mr. Gourley, will look for raised hands, and they will -- and will request you to unmute yourself for comment. When your comment is completed, Mr. Gorley will mute you again.

We do find it helpful if you identify yourself, but again, this is entirely voluntary, and you don't need to do so if you would rather not. Also, as a reminder, each person has three minutes. And please do -- please do keep your comments to three minutes or less. And accord -- and under the rules of the road, of Bagley-Keene -- the Bagley-Keene Open Meeting Act, comments are required to be tied to the agenda items. So any presentation today, please feel free to comment on that. Presentations from yesterday were appropriate to comment on yesterday.

Also, please realize that the Board and the speakers cannot generally respond to comments. But please, please do not take this to think that we are not listening or that we are being nonresponsive. It is important that we make sure that we comply with Bagley-Keene in order to avoid compromising either the commenter's goals or the Board's mission. So we are listening.

All information, including public comments, are being recorded and transcribed and will be available for the Board, for the agency staff, and indeed for the

public to review. And again, if you have any questions, please do write to info@cppa.ca.gov.

All right, with that, I hope that was clear. Are there any comments from members of the public?

MR. GOURLEY: Yes, there are, Chairperson Urban.

MS. URBAN: Please go ahead.

MR. GOURLEY: Ms. Loas (ph.), you may now unmute yourself. Thank you.

MS. LOAS: Hi. Thanks again for holding this informational session. I found it very helpful, just like yesterday. I did want to comment on automated decision-making, as that was the topic for today. So the CPRA -- actually, it's -- it's still quite -- it's not very clear whether the CPRA will provide an opt-out of profiling, so I think the CPPA is empowered to issue regs on that. But I think on that note, it might be helpful to understand whether, one, it'll -- it'll hinge on decisions that produce legal or similarly significant effects, and two, what those legal or similarly significant effects are. So just, like, providing some examples on that and whether we can take from the GDPR use cases, because there's a lot of resources on that end, or to what extent we can use those resources, as we operationalize some of these requirements.

So I think just kind of drawing the line of, you

know, what resources can we look to when we try to operationalize the opt-out of ADM, including profiling. Thanks.

MS. URBAN: Thank you very much, Ms. Loas.

Mr. Gourley, is there further public comment?

MR. GOURLEY: Yes, there is.

Ms. Huddleston (ph.), you may now unmute yourself.

MS. HUDDLESTON: Thank you. And thank you, Madam Chair, for hosting this informational session today, as well as the one yesterday. My name is Jennifer Huddleston, and I'm policy counsel with NetChoice. While it's worth acknowledging the concerns that were expressed by several of today's experts, I would ask that the agency should also be cautious about overly expansive actions that would penalize the use of neutral and beneficial technologies in a way that undermines their many daily uses that have benefitted conser -- consumers, including ways that technology such as algorithms improve and provide solutions for privacy, security, and authentication concerns.

As the Board considers their potential rulemakings on these issues, it should carefully consider the impacts on the beneficial uses as well as its attempts to address any concerns there are and also, as was mentioned earlier today, the impact on other issues, including speech, that

may arise from these regulations. With that in mind, the Board should focus on their roles on the authority they were given as it relates to privacy. Thank you.

MS. URBAN: Thank you very much, Ms. Huddleston.

Mr. Gourley, are there -- is there further public comment?

MR. GOURLEY: Yes, there is, one more.

Mr. Winters (ph.), you now have permission to unmute yourself.

MR. WINTERS: Hi. I'm Ben Winters (ph.). I am counsel for the Electronic Privacy Information Center. And I'd like to thank you for the opportunity to comment and for creating these public processes where we can get these great pub -- presentations. So I -- I pick plans (ph.) on writing. You know, we -- we commented earlier and planned on continuing to comment. But just in terms of automated decision-making systems, I'd like to urge the commission to -- or the -- the agency to adopt a broad rights-enhancing definition of automated decision-making technology, as well as profiling, ensure easy access to information about the use and logic of automated decision-making systems, and make it as easy as possible for individuals to opt-out of such systems.

That broad definition may, you know, be met with concerns from industry and even individuals with

beneficial uses of automated decision-making technologies, but that should be a burden that they can fulfill, and the risk of underinclusive definitions is -- is a -- is a greater one. And so we will provide more specific comments and suggestions to the agency throughout this process on how to define those. But those are substantial and important risks and -- and rights. And again, thank you for the opportunity.

MS. URBAN: Thank you very much, Mr. Winters.

Mr. Gourley, is there further public comment?

MR. GOURLEY: Yes, there's another one.

Mr. Winagle (ph.), you are now available to unmute yourself.

MR. WINAGLE: Great. Thank you. So I just wanted to make one quick con -- comment. I am counsel from Ultimate Kronos Group, and we deal a lot with AI on the employment side. So the -- the first general comment that I wanted to make is that, as we've seen in the EU AI proposal, there are significant differences between AI when it's used in different sectors. Now, obviously, when we're looking at general data protection regimes, they are trying to craft very broad solutions that -- that cut across all sectors. But in particular for AI, I think as we see that Europe and the EU is recognizing that this may not be kind of, like, a one-size-fits-all

type solution. So in the EU AI proposal, they're looking at when AI may affect -- may be high risk in certain areas.

And -- and one of them that they look at is -- is for example, employment. But -- but I would note that it is kind of important that when we look at how AI is going to work with respect to employment, we already have a number of regulations, as many of the speakers have -- have pointed out today. Discrimination is a serious issue when it comes to AI. But in employment, we already have, in California and in the United States, very strong anti-discrimination laws. So it's important to think about how these systems are going to interact with the existing laws and make sure that we are not essentially overregulating when we are putting certain laws in place for AI and employment.

That is, of course, assuming that the C -- CPRA is -- is actually going to apply to employment and -- and that -- that application doesn't go away for some point -- at -- at some point in the next year. That's it. Than -- and thank you for everyone, for listening.

MS. URBAN: Thank you very much. Very much appreciated.

Mr. Gourley, is there further public comment?

MR. GOURLEY: There is no comment at this time.

MS. URBAN: Thank you, Mr. Gourley. I will wait just a few seconds in case anybody's fiddling with the way they raise hand or is thinking.

MR. GOURLEY: There is one more.

MS. URBAN: Okay. Thank you.

MR. GOURLEY: Okay. Ms. Smith, you are now permitted to unmute yourself.

MS. SMITH: Hi, this is Nicole Smith again from ServiceNow. I just wanted to thank all the pre -- presenters for a wonderful presentation. It was very helpful. And I'm privacy counsel for ServiceNow in Silicon Valley. And as you know, Silicon Valley and California -- throughout California, we are a cradle of development for AI.

There's a lot already in development, and many companies like ServiceNow, Workday, Google have a set of AI principles already that they're holding themselves to in an effort to -- and to future proof the technology in anticipation of there being requirements to make sure that the AI isn't discriminatory, et cetera. And it looks like this is a direction that not only California but many other states and countries are going.

And to that end, it would be wonderful even -- if you could let us know in advance, so that we can adjust early on, rather than later in the development process,

and take into account any additional responsibilities that you see coming down the pipeline, essentially, as soon as possible, even perhaps before the final -- the rules are finalized, just because it takes a lot to be able to pivot. So the earlier we have the information, the earlier we can do a course correction, and it just would be greatly appreciated to get any kind of tips or outlines in advance, to the extent that it's available, even prior to the final rules being promulgated.

That's -- that's my comment. Many thanks.

MS. URBAN: Thank you very much, Ms. Smith. And thanks to everyone who has commented.

Mr. Gourley, are there further public comments?

MR. GOURLEY: There are no other commenters at this time.

MS. URBAN: Okay. Again, thank you to everyone who has provided such useful public comments over the last couple of days. And again, a big thank you to our presenters for their careful, informative, and rich presentations. They are very much appreciated.

Just a quick reminder that recordings -- a recording of the sessions and the presentations that speakers use will be on our website under meetings and events when they are processed. They do have to be put through processing in order to be accessible, but we hope that

they will be up soon. And when a transcript can be produced, that will also be put up.

I would be -- feel remiss if I didn't make one more plug for the stakeholder sessions while everyone is still here. Please do go to our website and check out the stakeholder sessions, and sign up if you are interested in doing that. We would greatly appreciate that.

And with that, I will move to our very last item, which is adjournment. Again, thank you to the presenters. Thank you to everyone who engaged in public comment. Thank you to board members and staff for all the work that you've done, and the -- and putting the meeting together, staff across several agencies, for all of their contributions to these informational sessions, and to the Board's work. With that, these informational sessions at the California Privacy Protection Agency are adjourned. Thanks, everyone.

(End of recording)

TRANSCRIBER'S CERTIFICATE

STATE OF CALIFORNIA

This is to certify that I transcribed the foregoing pages 1 to 128 to the best of my ability from an audio recording provided to me.

I have subscribed this certificate at Phoenix, Arizona, this 24th day of April, 2022.

Mariam Ayad

eScribers, LLC