PrivacyCon 2020 Part 1. July 21, 2020

ELISA JILLSON: Good morning. On behalf of my colleagues of the Federal Trade Commission, I'm happy to welcome you to our fourth annual PrivacyCon. My name is Elisa Jillson, and I'm an attorney in the division of Privacy and Identity Protection. My co-organizer for today's event is Jamie Hine, a senior attorney in the same division. Before we get started with our program, I need to review a few administrative details.

We're happy to welcome you via the webcast. We will make the webcast and the other workshop materials available online to create a lasting record for everyone interested in these issues. That will include links to the research discussed, and in a few weeks, a written transcript of today's event. As you may know, PrivacyCon is typically an in-person event. If there are any technological issues with this webcast, we will work to address them promptly, and we ask in advance for your patience if any such issues arise.

We'll be leaving time at the end of each panel to take questions from the audience. You can email your questions to privacycon@ftc.gov. If you would like to ask a question by Twitter, please tweet your question using @ftc and #privacycon20. Please understand that we may not be able to get all of the questions.

Lastly, I wanted to thank all of the researchers and the panelists for their participation in today's event. We are very grateful for your work in this important area. This program would not be possible without the great work done by many of our FTC colleagues. We'd like to thank our colleagues that assisted us in reviewing all of the research submissions, including Monique Einhorn and Patrick McAlvanah. We'd also like to thank those moderating the panels today, including Alan Conley, Phoebe Rouge, Daniel Wood, and Lerone Banks.

Finally, this conference would not be possible without the help of Crystal Peters, Marissa Henderson, James Maureen Bruce Jennings, paralegals Leah Singleton and Alex De Iglesias, June Cane from our Division of Consumer and Business Education, [INAUDIBLE] from of the FTC media team, Juliana Henderson and Nicole Drayton in our Office of Public Affairs, and Shawn Whitaker at OpenExchange. Thank you all. It is now my honor to welcome the director of the Bureau of Consumer Protection at the Federal Trade Commission, Andrew Smith.

ANDREW SMITH: Thank you, Elisa. Welcome to PrivacyCon 2020. Thank you, all, for being here virtually. This is the fifth year that we've held PrivacyCon, which brings together researchers from around the country and around the world to discuss cutting edge issues related to consumer privacy and security.

I know that you will all miss the opportunity to see each other face to face, but the most important feature of PrivacyCon remains the same, the spotlight on top notch research from a distinguished group of academics, physicians, economists, and other practitioners. Over the past few years privacy, PrivacyCon has been critical in keeping the FTC and other stakeholders up to date on emerging technologies and related privacy and data security risks.

PrivacyCon informs all of the work that we do here at the FTC, whether it be enforcement, business or consumer education, or rulemaking and policy efforts. In light of that influence, I'll start with a few words about what the FTC has been doing to protect consumers' privacy since the last PrivacyCon. Vigorous enforcement is at the heart of what the FTC does, and in the past year, we've brought privacy and security cases under the Fair Credit Reporting Act, the Children's Online Privacy Protection Act, the Gramm-Leach-Bliley Safeguards Rule, and our own FTC Act.

Shortly after last year's PrivacyCon, we announced settlements with Facebook, Equifax, and YouTube that shattered prior records for civil penalties or consumer redress for privacy and security violations. These settlements also required important structural changes with respect to how these companies treat consumers or children's information.

More recently, we brought a trio of cases against operators of mobile apps that failed to protect the privacy of children's information or misled consumers about compliance with children's privacy laws. The stalking app Retina-X, the Swiss mobile gaming company Miniclip, and the kid's app purveyor HyperBeard.

In recent data security cases, like Tapplock and Info Trax, we've put a stop to misrepresentations about smart locks security, and also to the failure to safeguard consumers' sensitive personal information. In privacy cases like on [INAUDIBLE] and Mount Diablo, we've challenged companies that made empty promises about keeping sensitive information, like financial information and emails, away from prying eyes.

We've also focused on educating business and consumers about data-related risks. For example, in recent months, we've issued guidance to businesses on how to develop coronavirus-related technologies that take privacy into account. We've offered advice on secure cloud computing and tips for using artificial intelligence and algorithms. For consumers, we put out guidance on how to safely use videoconferencing services and how to protect children's privacy while doing remote learning.

Rather than talking about past accomplishments, today's conversation needs to be focused on what the FTC should be doing going forward. Panelists today will discuss technologies ranging from mobile health and disaster apps to interconnected devices, such as smart speakers and cameras, to online ad delivery systems.

Economists will report on their studies abroad to gauge the effects of privacy legislation in Europe. And researchers will describe mechanisms for consumer choice and how consumers protect themselves from identity theft. The papers presented today will highlight technological developments that could be a boon to consumers, but that also present risks to privacy, security, and, in at least one instance, equal opportunity.

One final note before I turn the discussion over to our first panel. In our call for research papers, we specifically asked for research on mobile health apps, and the first panel of the day will be devoted to that important topic. Why health apps?

Industry reports show that consumers are increasingly using a variety of health-related apps, including fitness trackers, mood journals, smoking cessation or addiction aids, heart rate or sleep monitors, fertility trackers, diet guides, and more. Use of contact tracing apps during the COVID-19 pandemic could add a whole new dimension to that trend.

Earlier this year, the Department of Health and Human Services issued rules that will make it easier for consumers to access medical records through the app of their choice. This expanded access to health information could be an enormous benefit to consumers. But as we all know, wherever data flows increase, the opportunity for data compromise increases as well.

We, here at the FTC, have been active on health privacy issues, with cases like Practice Fusion, PaymentsMD, and Henry Schein. And we won't hesitate to take action when companies misrepresent what they're doing with consumers' health information, or otherwise put health data at undue risk.

Research like that presented today helps us to identify critical risks to consumers' health information, or other sensitive data, and better target our enforcement education and policy efforts. And so I want to thank all of the researchers who submitted their work to PrivacyCon, and all of the researchers who are presenting their work here today. What you do is of vital importance, and we look forward to hearing what you have to say.

And a big thank you to everyone who made today's event possible. I want to thank Jamie Hine and Elisa Jillson for leading the planning of this PrivacyCon, and also the many other FTC colleagues, from the Division of Privacy and Identity Protection, the Bureau of Economics, the Division of Business and Consumer Education, the Office of Public Affairs, and the Office of the Executive Director, who have worked together to make today's event possible. Finally, thank you to everyone who's attending virtually.

We appreciate the opportunity to engage with the public on this important and cutting edge research, and I hope that you enjoy the FTC's PrivacyCon. So our first panel begins at 9:20, and I'll turn it over to Ellen Connelly and Elisa Jillson for that panel. Thank you.

ELLEN CONNELLY: Good morning, everyone. Welcome to PrivacyCon 2020. I am Ellen Connelly, and my co-moderator today is Elisa Jillson. We are both attorneys in the Division of Privacy and Identity Protection at the FTC. We want to welcome you to our first panel of the day, which is entitled Health Apps. We have five panelists here to present some very interesting research.

First will be Quinn Grundy. Quinn is an assistant professor at the University of Toronto, and will present her research on the data sharing practices of medicines apps. Next we have Kenneth Mandl of Boston Children's Hospital and Harvard Medical School. He will present his paper on the privacy implications of moving health data, such as electronic health record information, to entities that are not covered by HIPAA.

Then we have Dena Mendelsohn of Elektra Labs, who will tell us about her work evaluating the privacy risks of connected sensor technologies in medicine. We will conclude the presentation

portion of our panel with John Torous and Sarah Lagan from Beth Israel Deaconess Medical Center and Harvard Medical School, describing their effort to develop a practical framework to aid consumers in their evaluation of health apps.

We have more detailed bios for all of our panelists available on the PrivacyCon website at ftc.gov. At the conclusion of the presentations, we'll have a question and answer period, during which we'll be able to have further discussion about the research presented. We'll be taking questions from the audience during the Q&A portion of the event, so please send your questions to privacycon@ftc.gov and we will try to include them. With that, I will turn it over to Quinn to start us off. Quinn?

QUINN GRUNDY: Thanks, Ellen, and thank you, again, for the opportunity to speak today. I am very excited to share with you some work my colleagues and I did looking at the data sharing practices of apps that have to do with medicines. And I'm hoping to spark some discussion this morning about how we think about data sharing within the context of the wider mobile ecosystem.

Could I have the disclosures slide, please? I first just wanted to acknowledge that this project was funded by the Sydney Policy Lab at the University of Sydney, and that we have no conflicts of interest. Next slide, please.

"Drug App Comes Free, Ads Included." So this was a headline that ran in The New York Times, back in 2011. This app, which is really popular among health professionals, provides information about prescribing, drug information, and clinical conditions.

This article reported, however that Epocrates was generating the bulk of its revenue from pharmaceutical companies that purchased targeted tailor advertising that was delivered to users on the basis of their personal characteristics and browsing history. So we know that apps routinely and legally share consumer data with third parties, and that this is done in exchange for services that aim to enhance the user's experience, such as integration with social media or to monetize the app.

But what about health apps? I think we all have a sense that health information is particularly sensitive, particularly personal, and also that it is valuable. We know that little transparency exists around data sharing, and also that threats to privacy are heightened when data are aggregated across multiple sources. But consumers are in a really difficult position and really have very little way of knowing whether their apps or websites that they use share this data and with whom.

So we wanted to add to this ongoing discussion by specifically examining the data sharing practices of a sample of apps that we thought were likely to share sensitive, specific health information that might be of high value to commercial stakeholders. So these are apps that provide information about medications, whether consumers taking medications or health professionals administering and prescribing. We wanted to know exactly what data these apps collected and where they sent it, and then to extrapolate from this data sharing to understand

where that data might travel beyond third parties within the wider mobile ecosystem. Next slide, please.

Our methods. So just quickly, and there is certainly more information provided in the paper we posted, but we looked at 100 paid and free apps-- sorry, we looked at the top 100 paid and free apps in the United States, UK, Australian, and Canadian Google Play stores, and screened it for any apps pertaining to medicine, so looking at the most popular apps.

We chose 24 of these apps that had some degree of interactivity. We designed a fake user profile, and in a lab setting, we interacted with these apps to simulate use. My colleague, Andrea Quintanilla, developed a tool, [? Adridento, ?] that performed a traffic analysis to eavesdrop on the data sharing that these apps performed between themselves and the network.

We analyzed the types of data shared and the IP addresses where it was sent. We were able to identify the entities that had these IP addresses, and then looked at their websites and these companies' privacy policies to understand what they might do with user data.

And frequently, we found that they reported further sharing through integrations or other commercial partners. And so we were then able to identify what we called fourth parties, and to simulate a worst case scenario of all the possible data sharing within this wider mobile ecosystem. Next slide, please.

So in a sample of just 24 apps, a tiny fraction of the health app market, we found that the majority did share user data outside the app with the network, and that some apps reported additional sharing within their privacy policies. We had pre-specified types of user data that might be shared, including names, time zones, medications, or email. Next slide, please.

We found that most commonly, apps we're sharing technical data, which might seem very benign on the face of it. So things like the device name, the operating system. What we did find that just over a third of these apps shared unique identifiers, such as Android IDs or email addresses. And a quarter shared the user's medication list, which is something that people could use to infer information about other sensitive things, like health conditions.

We conducted a network analysis of the data sharing relationships between the apps and these third parties. So we identified 55 unique entities that received or processed user data, which included app developers and their parent companies and these third parties. We found that third parties received a median of three different pieces, or unique transmissions of user data, and as many as 24 different types of user data.

In this network, you'll see the orange nodes are the apps, and the size of the node is the volume of user data sent or received. The blue nodes are third parties that we characterized as infrastructure, and represented about a third of the recipients. These were providers such as data storage, cloud providers. And because of their business model, which often involves keeping information secure, we reasoned that risks to privacy were low for this type of sharing.

The gray node,s however, are entities that were involved in the collection, collation, analysis, and then commercialization of user data, and this involved advertisers, social media, or analytics companies. And because of their business models and the way they described handling user data, we reasoned that there may be privacy risks associated with this type of data sharing. Next slide, please.

So first parties. We're calling first parties developers and parent companies that were receiving user data in our traffic analysis. We found that they received both the greatest volume and the greatest variety. And that might be expected, as this data was likely used to enhance the service that developer provided to users.

However, we also found, analyzing their websites and privacy policies, that developers were using this data for their own marketing purposes for products and services, but also the ability to tailor sponsored content, to sell advertising space, beyond banner ads, for example, and even to sell de-identified and aggregated data or analyses to third parties, like pharmaceutical companies or health insurance.

So for example, one app said they commercialized what they called the patient insights, from how medicines are used in the real world to health stakeholders, like pharmaceutical companies. And so the sense that because developers were collecting information, that that might be safe and secure and private, may not, in fact, be entirely true. Next slide, please.

When we looked deeper into the third parties receiving user data, there were 21 entities that we characterized as analytics. We found, when we analyzed their privacy policies, that these entities typically reserved the right to collect de-identified and aggregated data from app users for their own commercial purposes, and to share these data among their commercial partners, or to transfer data as a business asset in the event of a sale.

What was interesting was that for third parties, their privacy policies defined a relationship with the app developer, not the app user. And so if app users were concerned about the collection or sharing of their data, even if it was de-identified or aggregated, they were referred back to the developer in the event of a privacy complaint and couldn't take it up with the third party directly. Next slide, please.

So fourth parties. We found that the third party entities reported this ability to share end user data with 216 different fourth parties, so entities beyond what directly received user data. And we found that these entities could potentially create highly detailed profiles of users, even if they could not identify them by name.

So while certain data sources are clearly sensitive and personal, or identifying, like your date of birth or a drug list, others may seem irrelevant from a privacy perspective. However, when combined, all these little pieces of information from a variety of different sources can create a fairly detailed picture of a user or to associate them with certain groups. So we conducted a network analysis to understand, again, how data might be aggregated within larger companies and their commercial partners, and we simulated this hypothetical data sharing. Next slide, please.

So this very busy picture is our fourth party network, and it's the worst case scenario, where, if all the data were shared by all these apps within the network, 44% of these fourth party entities may have access to medical information, and all but four of them also had access to potentially identifying personal data.

We found that multinational technology companies, digital analytics, and advertising firms occupied highly central and prominent positions within this data sharing network, with a significant ability to aggregate and potentially re-identify users. And of interest, only 1% of these entities could be considered as health-related or part of the health sector. Next slide, please.

So in discussion, I think what these results suggest is that collection and commercialization of health app users' data is a legitimate business practice, and that sharing of user data is both routine and far from transparent. Our analysis suggests that privacy regulation must emphasize the accountabilities of both those that collect and control user data-- right now, a great deal of onus is placed on developers-- but also that process it, these third and fourth parties that sit behind the scene.

I think we increasingly understand that the sharing of app user data ultimately has real world consequences. And I think the panelists in later talks today will be sharing some of these, things like bias in algorithms. These consequences include highly targeted advertising or the commercialization of data into algorithms that ultimately make decisions about people's insurance premiums, employability, or financial services.

We're seeing increased scrutiny of collection and sharing of sensitive, personal, or health data, but I think understanding how data are aggregated suggests that in combination, a much wider array of data types might actually be considered health data and used to make inferences about people and groups. So for example, even the existence of a health app or a mental health app on one's phone could be used to make inferences and decisions about a person.

Our current regulation focuses on securing individual informed consent through improving privacy policies or labels for apps and protecting harms to individuals, for example, by ensuring that data are de-identified. However, when we think about the mobile ecosystem, the aggregation and sharing of data within this wider space, I think we also need to consider the disproportionate harms that can occur to certain groups when inferences are made on the basis of characteristic. Next slide, please.

So in conclusion, I wanted to share our dashboard, healthprivacy.info, where the full data from this study are available, and it includes additional information about the security analysis we also performed and the apps that we sampled. I'd like to thank Ellen and Elisa, again, for this opportunity, and to acknowledge my collaborators on this project, and in particular Andrea Quintanilla, for developing the tool we used in the traffic analysis. And I'd like to thank, again, the Sydney Policy Lab and the Australian Communications Consumer Action Network, who we worked with. Thank you so much.

ELLEN CONNELLY: Thank you so much, Quinn, for that really interesting presentation. We're going to move on now to our next presenter, and next we'll hear from Ken. Ken, you're up.

KENNETH MANDL: Terrific. I'd like to thank the FTC organizers of PrivacyCon for putting together this spectacular program, and I'm honored to be able to participate. Let me set the context for my talk.

At the beginning of the Obama administration-- and I assume my slides are going up-- at the beginning of the Obama administration, Congress passed the HITECH Act, and the federal government invested $48 billion to promote the adoption of electronic health records. Because I had worked with electronic health records as a physician and a researcher, I knew that these older 1980s and 1990s software stacks would not advance the goals of a learning health system, where the data collected are put to work to improve health, control costs, drive discovery, underpin public health, and empower patients to manage their care and participate in research.

So I wrote in The New England Journal of Medicine a piece proposing that if we're going to invest this $48 billion of federal dollars-- which, by the way, was complemented by probably between a half a trillion and a trillion dollars of private and public investment in installing these electronic medical record systems and purchasing them-- if we're going to do that, why don't we think about a public interface that essentially turns the electronic health record into a smartphone-like platform that can run apps that can be added or deleted the same way they could on the iPhone?

And when we wrote this, the iPhone was one year old, and we were just starting to see the power of an application programming interface that allowed third party apps to connect to a platform. The type of business advances, the types of innovation, the competition that you see in an app store, the truly spectacular examples of apps that were emerging, could we have this for medicine, too, even though we were investing in older technology as the backbone of our health IT infrastructure?

So we were funded for $15 million by the Office of the National Coordinator. And what we proposed was an application programming interface that would enable EHRs to run these apps. This was a high risk play, because each EHR was different, had no standard for the storage of data, and was not designed to ever let data out of its walls.

In fact, quite the opposite. Patients had some access to their electronic health record through portals. Many of you may have used them. But those data are essentially behind glass. You can look at them, but you can't get a computable copy. You can't feed them into a computable process, like an app or an algorithm.

Now, HIPAA, passed in 1996, guaranteed that consumers could get access to a copy of their data in an electronic format if it was feasible. And from 1996 until, essentially, a year or two ago, it was determined by health care and health care IT vendors that, in fact, it was not feasible.

Now, whether that's true, I think, is a subject of debate. But the good news is that now, 10 years after the $48 billion investment began, we have actually new regulation that comes from the Office of the National Coordinator of Health Information Technology, an HHS agency that oversaw the $48 billion investment and that funded us and that now has passed regulation based on the 21st Century Cures Act.

I don't do very much lobbying, but I managed to get this one sentence into the 21st Century Cures Act, requiring an API that provides access to all data elements of a patient's electronic health record, and that those elements can be accessed without special effort. This underpins the potential for an extremely robust apps economy.

A second API was also developed in our group and managed to make it in under the wire into the regulation, which allows us to get data on populations out of electronic health records as well. The first API is called Smart on FHIR. Next slide, please.

And these two APIs together allow us to potentially think about the health innovation in a parallel way to how Tim Berners-Lee thought about the web. I think the slides might be a bit ahead. There should be a slide of Tim Berners Lee showing now on the worldwide web.

In a sense, what we're trying to do for health care is similar to what he tried to do. He wanted to share pre-prints of his articles, and he invented a way to show those articles in HTML. He invented a web server so that you could serve up those documents. He invented HTTP so that you could link to them, and he invented a web browser so you could display them.

What Tim Berners Lee created parsimoniously, and then instantiated through the world wide web consortium, enabled a tremendous economy to be built on top of these parsimonious rules and specifications. The API is regulated by the Office of the National Coordinator, stemming from the 21st Century Cures Act, actually have the potential to create innovation within the health care domain. The next slide should have a picture of the Apple Health app with the heart on it.

And the first major company to take advantage of these APIs, even before these final regulations, based on some earlier regulations, was Apple. And Apple had a spectacular success. They used our API, called Smart on FHIR, to connect the health app to hundreds of health systems so that patients at all those health systems could download data from the health system onto their phone and expose it to other apps.

And there it is. There's the API and the health app being announced on the Apple stage. To the right of the health app you see this little blue button, 2.0. This is less well known, but it's actually a very important effort, met by CMS, to enable all consumers to have access to their claims data through the same Smart on FHIR API. And as I mentioned, though not the subject of the talk today, a second API, called Bulk FHIR Access, is going to give us data on whole populations.

The next slide has a picture of the USCDI. The data that we're talking about is regulated as the United States core data set for interoperability and defines which health system data will be available through these APIs. This data set will expand over time, but now includes things like medications, diagnoses, laboratories.

The next slide shows the data protected by HIPAA on the left and the smart API in the middle, where the patient can request the data, for example, to be downloaded into their Apple Health app. And then the magic that happens here is that the patient gets a copy of their data.

The regulatory, piece which has not been fully addressed, is that the data goes from HIPAA covered, in the health system, to FTC covered afterwards. And what happens as the data are passing across the API is critical for protection. The FDA has the most enforcement power over privacy in the US, but it does not prescribe what those privacy requirements are.

The next slide shows some aspects of privacy policies that are in the rule, that they be written in plain language, that they be made public accessible all the time, that they include statements of whether and how the data is accessed, used, or sold, that they share this with users before accessing the data, and that they require express consent. So it establishes some elements of what needs to go in a privacy policy, and that is a good start.

The next slide, Analysis of Current Approaches, shows us that, yes, there are a few community-based efforts to address this. There is a model privacy notice. There are questionnaires that some of the electronic health record companies have actually developed to ask app developers what their intentions are. There are external codes of conduct. An early one comes out of something called the CARIN Alliance, and it gives us an attestation that is enforceable later, by the FTC, as to what that company will do with data collected by the app.

The next slide shows that there was opposition to this rule on the basis of multiple special interests. I strongly supported the rule publicly, but I have to agree with one of the points that was made in the opposition to the rule. And the rule was passed over this opposition, and I'm going to talk about some approaches that we're taking to address the point.

The point is that when data traverses that API, it loses, potentially, a lot of protection. And the opportunity here is to enable the FTC to handle the proper stewardship of those data. I addressed some of these points about the privacy of data once it has traversed the API and lost the HIPAA protections, in The New England Journal, around what do we need to do to be data citizens in the 21st century?

We have to be very cognizant that there will be, as an exception to the rule, I'm sure, but nonetheless, predatory app companies. We may have multiple forces, partially driven by privacy concerns, where we don't get the market economy of apps competing with each other and adding value to the health system.

If we're not careful about the security, we'll have abuses and breaches, which will lose confidence. And also, we must be very careful about widening the digital divide when we deal with these technologies and when we deal with people's attitudes towards privacy, which may, actually, vary across this digital divide.

So I want to talk briefly about how there is a stop gap technical fix that is enabled while we think further about how to strengthen the FTC's role. And what that is-- and what we should do now is go to the slide that shows the smart app privacy manifest, which is a couple of slides down.

And the opportunity here is the following. The API provisions were accompanied by very strong regulations against information blocking, so that a health system cannot prevent a patient from

choosing an app that they wish to connect to their electronic health record. An electronic health record vendor cannot prevent a patient from connecting an app.

Overall, that's very good, because it gives patients agency, and it gives app developers and innovators the opportunity to have a large market. The problem is that it could be perceived of as information blocking, just to tell patients and warn them about bad apps, because bad apps may be in the eye of the beholder.

And so the Office of the National Coordinator, in the regulation, actually addressed this with a potentially innovative solution. And that is that in the OAuth process that enables the authorization and authentication of the user and the app to the electronic health record, there is an opportunity to present the manifest of privacy policies. And in fact, some of the electronic medical record companies have begun to do this.

And so there is, specifically regulated, an approach that this will not be information blocking if basic information is provided. What kind of information could we provide? The location of the privacy policy, the data storage policy, the data usage policy, the data sharing policy.

Who made the app developers send data to and for what purpose? What relevant data? The apps method for approaching patients before sharing their data with other parties, as we heard about from Quinn. And we can also put in trust entities badges, if the apps have actually attested to certain practices.

However, what we may also want to be sure that we do is to also-- and we can go to the last slide, which is this timeline-- is make sure that this decade of work that has gone into liberating information from electronic health records to empower consumers and provide them with computable copies of data actually results in a safe ecosystem.

Part of this is defining what the privacy policies are and making sure, perhaps even from a regulatory perspective, that those elements are there. Research is needed on how patients understand those privacy policies, and I believe the FTC could have a strengthened role in enforcement of those policies, as well, to make sure that when there are breaches of what is promised, that there is a strong enforcement reaction.

And it's very critical to protect consumers from harms related to health data. And if we can make consumers feel safe in this environment, I think the opportunity is almost unlimited. Thank you very much.

ELLEN CONNELLY: Thank you so much, Ken. Dena, you're up next.

DENA MENDELSOHN: Hi. My name is Dena Mendelsohn. I'm the Director of Health Policy and Data Governance at Elektra Labs. We offer services to better evaluate and dispense connected health moderating technology, many of which feed into the health apps that you're hearing about today.

Prior to joining Elektra earlier this year, I served as senior policy counsel at Consumer Report, where one of my most recent projects was reviewing the data practices and security of a handful of reproductive health apps. Today, I will discuss a paper published by my colleague, titled "Modernizing and Designing Evaluation Frameworks for Connected Sensor Technologies in Medicine." Next slide, please.

In today's presentation, I'm shifting gears slightly from the preceding speakers, and will pan out to consider the ecosystem that feeds into and works with health apps. I will give you a broad overview of why clinicians are increasingly using biometric monitoring technologies, and what type of due diligence we recommend before adopting this remote monitoring technology. I will conclude with what we recommend to simplify the decision process. Sneak preview, it's a label, somewhat akin to a nutrition label that we're all familiar with. Next slide, please.

But first, let's talk about why, why collect digital measurements in real time at home. Well, the simple answer is that in researching care, remote sensing offered a more holistic view of a person's lived experience, especially when we're looking at chronic conditions that impact the person's daily life.

Do we want to just know how they're doing through a few status points throughout their day? Well, not really when there's a better alternative, where we know how they're doing continuously throughout the day and over a longer period of time. So while it would be simple to just step away from health apps, for those who are concerned about their data rights, that really takes away some very powerful tools for them, and so it's not what I think any of us would recommend. Next slide, please.

We believe in the value of the remote health monitoring technology before COVID-19 took over, but the value of this technology is even more clear during this difficult time. Uptake of remote monitoring technology, like connected sensors, are likely to rapidly increase during this pandemic, especially following guidance from the FDA and CMS that encourage widespread use. I think we've all seen a lot of articles about this in the lay press. Yet, public discussions of the risk of these technologies has been limited. Next slide, please. We should be at the Due Diligence as Necessary slide.

And this is where, in our paper, we provide a deep dive into the due diligence that is critical when selecting connected sensor technology, whether it feeds into a health app or not. Next slide, please.

What you're seeing here is a broad overview of our five-point holistic framework for balancing the benefits and risk of adopting connected health technology. Again, many of this technology feeds into the health apps that we're talking about in this panel. The first three dimensions evaluates the data and subsequent results generated by connected biometric monitoring products.

The fourth dimension, utility and usability, evaluate the ease of implementation and adoption of the product. And the last dimension, economic feasibility, has the reader consider the cost and the value of adoption. As explained in the paper, evaluations should be multidimensional, and a single score should be avoided. Next slide, please.

So on this slide, we're looking at step 1 of the evaluation framework. And this is less about health apps and more about ensuring that the technology that's being used will generate information about a user that is suitable, both in terms of what measurements are made, the accuracy, and the appropriateness in the situation where it will be implemented. Next slide, please.

As discussed in the paper, suitable technology must be verified and validated. Simply put, the technology must be accurate, both in the measurements it makes, as well as any algorithms that applies to the collected data, and that the technology were for a specific use case in mind. After all, not all technology is appropriate in all contexts. Next slide, please.

The second part of the evaluation framework in this paper considers security. Next slide.

On the Cybersecurity Considerations slide, the paper recommends including whether the company has a coordinated vulnerability disclosure policy and what's in it. Does the organization [? public ?] a security support lifetime and issue secure, prompt, and agile software updates once security issues are discovered? And finally, does the organization track and share a software bill of materials? Next slide, please.

A third component, and probably of a special interest to viewers today, is to look under the hood of data rights and governance. Given that you're streaming PrivacyCon, you probably know why data rights are an important safety tool for users of technology. When it comes to technology involved in health and health care, individuals' right to data governance is pretty uncertain. Next slide, please.

As it is, in our health care system, we have strong protections for patient bio specimens, like blood or genomic data, but protections are murkier for digital specimens. The same can be said of data created by health apps. Make no mistake, wearables, health apps, and in-home sensors offer great promise for affordable, accessible, equitable, high quality care. But in the modern era, data rights have become a safety issue that extends beyond the body. The digital health data that folks generate may threaten both their health and their financial welfare, which you're hearing a lot about today. Next slide, please.

We've seen enough headlines to know that there's a problem with how data is collected, used, and shared. Next slide, please.

[? Shaky ?] data rights in the United States means that when clinicians recommend some health technologies to their patients, or friend recommends it to another friend, they could be unwittingly putting the individual at risk. That's why the third part of the evaluation framework asks these foundational questions about the data practices of technology under consideration.

As explained in the paper, there could be gradations in manufacturer data practices. In our evaluation framework, the minimum threshold is that the manufacturer has a EULA, or terms of service, and privacy policies that are publicly accessible online. But really, we know that that's just a baseline. It's also important that documents are comprehensible, or understandable, by a broad audience.

And at the end of the day, being fully transparent about practices is not the final solution. Transparency is not the solution, but rather, manufacturer and app developers need to commit to privacy protective practices. As we explained in the paper, the highest quality data practices means that the EULA and terms of service do not contain exculpatory language. There should also be an opt-in or opt-out of third party transfer or use of data, where appropriate. And ideally, this right should remain unchanged, even in the case of a change in ownership of the connected technology or the sensor manufacturer. Next slide, please.

Finally, parts 4 and 5 of our framework consider whether a product has features that users need and whether it's designed in a way that folks will actually want to use it. And finally, no evaluation will be completed without consideration of the cost and value of the technology. Next slide, please. We should be looking at the Nutrition Label slide.

Now that I've considered the holistic evaluation framework, I'll remind you that excellence in one dimension does not necessarily imply excellence in another. Indeed, significant deficiencies in any one dimension may lead to problems when using connected sensor technologies in research or in practice. Thus, we propose a framework that simplifies the evaluation process of connected sensor technologies for the intended use, but it did not give an individual score that would make a decision for the reader.

As remote health monitoring technologies becomes increasingly commonplace, more and more people need to decide the risk benefit type of evaluation that we explained in the paper. But this analysis will need to be more straightforward. As the paper concludes, they propose that a connected sensor technology label could be a useful piece of infrastructure for an evaluation framework, which would make it easier for decision makers to understand critical aspects of technology in a streamlined and accessible format.

It's extremely likely that remote health monitoring technologies, paired with health apps, and some connected in other ways, will become a very common thread in how individuals manage their own health, how health care is provided, and in the context of biomedical research. I would encourage viewers to read the paper that I discussed today to get a deeper understanding of the features of connected sensor technologies and their benefits and risks, and how they should be evaluated ahead of deployment.

If viewers from the health care sector are interested in learning more about digital medicine to enhance public health, I would encourage them to check out the Digital Medicine Society, or DiMe, which is a professional society for digital medicine. I also want to acknowledge the authors of this paper, my colleagues Andy Coravos, as well as Megan Doerr, Jennifer Goldsack, Christine Manta, Mark Shervey, Beau Woods, and Bill Wood. I also want to thank the FTC for inviting me to speak today and for its efforts in moving PrivacyCon online this year. Thank you.

ELLEN CONNELLY: Thank you so much, Dena, for that really interesting presentation. And now we'll move to our final presenters for this part of the panel. Our final presenters are John and Sarah. So John and Sarah, I'll turn it over to you.

JOHN TOROUS: Oh, thank you for having us, and as going forth, I think you'll hear some themes that are repeating and some parts that are new. But we'll start with the first slide. We'll see if it gets pulled up, Actionable App Evaluation. And let's see, is it up? I think it's not up yet.

ELLEN CONNELLY: We're experiencing a little bit of a time delay with certain browsers on the slides, so if you could maybe just start off, and hopefully they'll catch up pretty quickly.

JOHN TOROUS: So as I said, we'll talk about actionable health app evaluations. And first, we want to thank our donor, the Argosy Foundation, which made this work possible. We couldn't really have done any of this without their support.

And I think what we're talking about today-- and I think Sarah and I are coming from an interesting position, where we're doing clinical research, but we're also delivering clinical care. So we're looking at how these apps work in real world settings, and how policies really impact care decisions and patients today on the ground. And we know from experience there's many good smartphone health apps and wearables that can improve care. As we've heard about from other speakers, there's also some pretty concerning dangerous ones that can directly harm care, threaten care, or harm the whole field.

And we know, again, that a lot of these health care apps wearables are pretty clever i health and wellness devices. They don't really go under the medical category, so they work hard to avoid different types of regulation. So looking at the slides of privacy concerns, again, we know that many of these things live outside of HIPAA and other privacy laws.

And we know that when a lot of patients come to see us, they actually expect, when they go on to the commercial marketplaces, download an app or a wearable, that if it's related to health and they see things about health, they intuitively expect that it's going to offer health protection. So do many of our physicians, therapists, psychologists, social work colleagues, as well, and nurse practitioner nurses.

And again, that set a line between how is it regulated, where is the data going? And on the Privacy Concerns slide, you can see the same thing that Quinn Grundy presented, in that you don't always know where your data is going. And on the second Privacy Concerns slide, you can see our team did a paper last year, where we actually did something called a man in the middle attack, and looked at where was data from popular mental health apps, popular apps for depression and smoking. If you downloaded them, where was your data going?

And the trick was we actually did read those long, complex privacy policies, and what we've found is even if the privacy policy promised you and pinky swore that your data was really going to stay secure and safe and it wasn't going to go anywhere, it still went somewhere. Often, it went to Facebook Analytics, among other sources.

So even if the app developers did have a privacy policy, sometimes it wasn't actually followed, as well, which was pretty concerning. And that slide, you can see The Washington Post covered. The article saying, "Smoking and Depression Apps are Selling Your Data," which was a little bit concerning.

And certainly, these privacy concerns we've heard are still with us today. This is just a headline from February 2020, so not that long ago, about a popular therapy app that's disclosing different aspects of users' data. I think in mental health, we're in a unique position, that a lot of digital health actually focuses on mental health because we can both collect data from sensors and apps and forums care. And in mental health, we can also offer people treatments via videos and technology. So a lot of this is actually happening in the mental health space, and privacy concerns have actually shown up a lot in the mental health space, as well as other spaces as well.

You can see on this slide it has exaggerated claims of effectiveness. In a different study with a group led by the Black Dog Institute in Australia, we actually read the app stores and said, what are these apps claiming? If I'm a patient, I'm a clinician, I'm a position, I'm an NP and I'm looking at these apps, if you read the app store claims, that really, over half of them make claims that could be seen as medical, implying effectiveness.

We actually went back and tried to tie it down to what is actually claimed in the literature? What is actually proven? And really, it's less than 2%. So there's a huge dichotomy between what a consumer is seeing and what is actually supported. And I think there's different consequences we've heard different speakers through this misinformation.

On the Perils of Misinformation slide, one really concerning aspect we saw was that a lot of mental health apps just aren't updated. The developers aren't keeping them current, and some of these apps are offering incorrect suicide hotlines. And I think the quote speaks for itself, "Non-existent or inaccurate suicide crisis hotline phone numbers were provided by mental health apps, downloaded more than two million times."

So again, I don't think anyone's trying to give incorrect or false information, but again, sometimes these things are just not really able to live up to the goals and standards that they would want to. I think a lot of times, the way that people find apps, be it again colleagues that we work on in the hospital, patients, the people we talk to, is they look at, well, what's the top out there in terms of the search? Or which one has five stars? Or which one has over 100,000 downloads?

And that's not always the best approach to do it. If you type in schizophrenia, this app that's really a pawn game shows up. It's stigmatizing. It's incorrect. It actually doesn't work on a lot of phones, and that's probably a good thing. But again, just because it shows up highly in a commercial marketplace really isn't going to tell you a lot about the app.

Prior research, on the left, clusters, really shows you that even apps that have high star ratings, it doesn't really tell you much about your clinical utility or validity. And this was more than mental health. This looked at apps and diabetes, heart disease, as well as depression. And that kind of hockey stick graph, that sharp decline on the right, where the slides with stars and download metrics are misleading, shows you that, really, the average person who downloads one of these apps, they don't actually use-- about 95% of people aren't going to be using after two weeks.

The engagement really drops off. So even if the app is highly downloaded, the real question is, can people actually stick to it? You can't really learn that from metrics. So I'm going to have Sarah take over on this slide that says Deriving a Practical App Evaluation Framework.

SARAH LAGAN: So in light of these concerns, we're now going to briefly discuss our efforts with the American Psychiatric Association to develop a framework specifically for the assessment of mental health apps, but applicable to health apps broadly as well. So on the next slide, you'll see how there are numerous app evaluation teams. So there's a clear need for an evaluation system beyond app store metrics, as we saw with these [INAUDIBLE] concerns.

And to respond to this need, there's been numerous app evaluation frameworks that have emerged, including the NHS in England, Denmark's MindApp system, and over 45, as of 2018, with far more emerging in the two years since then.

So if we go to the next slide, the Potential for Harm with Lists and Static Ratings, many of these frameworks rely on lists or static ratings, which may fail to account for nuance and diverse app needs. Just as there's no A plus medication or talk therapy, people react to and use apps differently. Even the same app may be used in different ways, depending on individual variation and preference and needs. Further, the app market is constantly changing and very dynamic, and it's hard to know if these lists respond to the most current version of the app.

So if we go to the next slide, what we did was we looked at 45 different frameworks, back in 2013, and we sorted the 604 unique questions from those frameworks into categories. So as you can see on this graph on the right, short-term usability questions were highly overrepresented, compared to questions regarding privacy. The privacy questions are the ones in pink. So you can see on the right-hand graph how usability questions were just far more predominant, even despite the privacy concerns that you've heard raised throughout the presentation today.

If you go to the next slide, we use these questions to inform the framework we created with the APA a few years back. And as you can see this pyramid graph here, there are five levels, Accessibility, Privacy, Critical Foundation, Engagement, and Therapeutic Goal. Corresponding to each of these categories is an ethical principle. So our framework is really grounded in the ethics that guide care.

And in the years since it has emerged, if you go to the next slide, we'll see how it stacks up really well on privacy questions specifically. So this recent scoping review on different evaluation systems for apps features the APA model.

And you can see, highlighted on the left-hand side, how the APA model they found to be extremely sterile in addressing the various components of privacy. So data is being collected? To whom is it shared? And on the right-hand side, you'll see how the app has been pretty widely cited and mentioned in the literature since 2016.

So then you can go to the next slide a Framework to be Customized and Adapted. It's been referenced in numerous different papers, highlighting its adaptability beyond just mental health apps and towards health apps more prominently. So our next question this year was, how could

we use this framework and make it even more actionable for consumers, clinicians, patients, and any user of apps?

JOHN TOROUS: So what we wanted to work on was saying, well, we've built these principles. We've kind of guided people on what to look for, but that can put some more onus on the patient on the clinician in the visit. And again, we talked about, so how can we make it easier for people to understand it?

And one of the problems was, again, a lot of app rating systems will say, was is it easy to use? But again, what does ease of use mean? Who is it for? And if we say an app is easy to use, really, that's putting a value judgment on different people, trying to say what it is.

So we broke down things like ease of use into things like engagement style. Does it have peer support? Is it AI drip? And does it have videos? Does it have gamification? The idea isn't to judge it, but we wanted to make our criteria with different elements that could be objectively reproduced, so it's kind of yes, no, or numbers.

So we kind of smushed that APA pyramid into over 100 questions, which are more objective, to help people understand what an app could or couldn't do, what it offers. And again, the goal isn't to offer judgment. It's to say, what features or what elements does it have or not have?

So the idea is to build a system powered by the community. This was a theoretical model that we published last year. And right now we're at the A. We're looking at clinicians and patients using it, giving us be back. And the goal is to get more towards B, where we do get our developers involved as well.

But you can actually see our project live today at apps.digitalpsych.org, and you can actually use it. I think what I'm going to show you guys is this is what the broad database looks like. Again, we try to be fully transparent, so this is a screenshot of the website. You can see what it looks like today. But the idea is you can imagine this screenshot app actually be very, very wide because it's going to go for 105 questions of different apps.

And people can [INAUDIBLE] them. And people say, hey, what are all the apps? Again, [INAUDIBLE] in Spanish to have really great privacy features, and we'll show those. Or someone may say, I don't actually care about privacy and I want to find all the apps that have video, and I don't care. And the idea is we want to make people aware. We want to make sure people make informed choices, but we don't want to force what people's choices are.

As other speakers have said, we want to make sure people can pick what that nutrition label is. They have to be aware of all the information, but we're not here to say, this is the best one for you. Someone may say, look, it's very important my app have text messaging, and that's the most important thing and other features don't matter.

So people can easily search our database and learn about what features are in an app. There was recently an article in STAT news last week, kind of showing how we can help people make apps.

But the real question is, does [INAUDIBLE] actually change clinical decision-making, change impact?

And one thing we can do is because we have this database, we can query it. So one question we can ask of the apps we've looked at, do app support download? Do they offer more privacy features? As you showed the code in that line, but the answer was no. In contrast to last year's PrivacyCon, where the apps we've looked at so far, we said, do apps that cost more offer more privacy features? If you pay more, do you get better privacy? And again, from the subset of apps we've looked at, the answer was no.

We can also use this to help patients make smarter decisions. We can do patients with training. This was an app we don't endorse, or not endorse any app. But before, we asked a group of patients, would you be interested in downloading this app? And basically, it was 50/50. And after we had patients use a tool and ask questions, you can see that their decision-making changed. People said no.

We can also do this with clinicians. Again, just an example. Blue was before and then orange was after. We took a lot of clinicians who were in that three middle range. Some said, hey, I'm not as interested in this app now. So it's possible to quickly let people search for apps, learn, and change how they're making decisions. So we're expanding on those.

And I'd say that, certainly, I think clinicians and patients both are pretty excited to learn about this stuff, they just don't always consider it because they think that these protections are inherent. So we'll close by, again, thanking our donor who made this work possible, and the FTC for inviting us.

ELLEN CONNELLY: Thank you so much, John and Sarah, and thank you to all the other panelists as well. We'd like to move on now to our Q&A portion of the panel, and hopefully engage in some good discussion, expanding upon some of the ideas that you've mentioned and maybe touching on some new ideas. So I'll start us off, and I'd like to start with a question or two that are probably at the top of everyone's mind these days, and these are questions related to the pandemic.

So as you've probably seen, there have been a multitude of recent news articles regarding, for instance, a pandemic-induced mental health crisis in the US and a significant increase in consumer demand for things like therapy apps during the pandemic. Are there practical steps that a consumer can take now to protect her privacy while also obtaining useful health-related supportive services?

So John and Sarah, you've, in particular, touched on some of these issues in your presentation, so I'd like to maybe start the discussion by giving you a chance to expand upon this particular part of your work, and then I'll move on to the other panelists.

JOHN TOROUS: It's a very good question. In the pandemic, as people are looking for more mental health resources, how can they find useful ones and not end up, perhaps, trading all their

information in? We've seen this clinically in the patients that we support. People do want extra help.

And I think basically, what we always do with people first is we check for a privacy policy. You'd be surprised how many apps don't even offer the level one that even talks about even a privacy policy. But usually, what we actually do with patients is we look at how much information the app may be wanting, if it wants GPS for different levels.

And then what we do is say, what is the risk benefit? Usually, by talking with patients, people, again, are usually shocked that the app is collecting this much data, but then sometimes, oftentimes, they say it's not worth the benefit, but [? it is. ?] But I think as long as people are informed and aware, that's a very good first step, where then people realize the risk benefit and going through that.

Usually, as people bring apps to us, we're adding them to our database and then going over it with patients, and sometimes we use our database. If an app doesn't come up with a good match, patients will say, well, what if I was willing to compromise on this, or if I wanted more privacy? So usually, we have a discussion around it and it turns out to be, I would almost say, therapeutic and informative for all parties.

ELLEN CONNELLY: Thank you. I'd like to see if anyone else has anything to add. Maybe Quinn, do you have anything to add? Or Dena?

DENA MENDELSOHN: Yeah, I'll just add in the first step is making sure that individuals understand that, in many cases, HIPAA doesn't apply. So as speakers said a few times today, there seems to be some misunderstanding or assumption that when we're talking about health, that all health is protect the same, and it's just simply not.

And then going from there, just reminding consumers that health apps is a very large market. So there are choices. It's not that you always have to give up your data. You need to be careful about picking which one you're going to go with and just be intentional about your selection, rather than simply downloading the most popular app or the one that one person may have recommended.

ELLEN CONNELLY: Thank you, Dena. Quinn or Ken, do you have anything to add?

QUINN GRUNDY: Yeah, I might offer a slightly different perspective. I think the pandemic has laid bare, in many areas of our lives, pre-existing problems and really exacerbated them. And so I think this is a great example, where there's actually maybe greater awareness around privacy and security of data than ever before. And I think what that will hopefully lead to is some collective demand that there be better protections.

And I can't really think of another consumer sector or industry or product where the same amount of responsibility is placed on consumers for ensuring that products are safe to use. And I think as we learn more and more about the consequences of lack of privacy or privacy breaches, that hopefully, we will see some better regulation.

And an example would be there's no regulation, for example, placed on the app stores or app distributors to ensure that the products they market are safe for use, and we don't see that in other sectors. So while I think there are some practical steps, and consumers are in a position where they have to make choices for themselves, I don't think that, ultimately, it should be a consumer's responsibility to make sure that products are safe and private.

ELLEN CONNELLY: Ken, did you have anything to add?

KENNETH MANDL: I'll just add that, yeah, it's definitely the Wild West. I think one thing a consumer can do is to look for endorsements by professional organizations that they trust. Hopefully those professional organizations are educated on the issues we're talking about today, enough to know what to endorse. It won't always be the case.

And the other caveat, unfortunately-- because I'm sure many of these apps are very useful-- is that privacy policies and terms of use can change, including for the data that you've already contributed. And so I think we really do need stronger protections going forward so that consumers can take advantage of this emerging apps economy.

One advantage in these API-based apps, where we have the transition that I talked about from a HIPAA-covered entity to the FTC regulation, is there, we really know what the data going in are and we have the opportunity to regulate those data as they go into FTC jurisdiction.

With a mental health app, where it's really health-related but not coming from the health system, I think the oversight of those is even more complex, or as complex as it is to regulate the health API-based apps. Regulating apps that provide a health benefit is, I think, even more complex, but comprehensive legislation is probably what we need.

ELLEN CONNELLY: OK, Dena, I see a hand raised, and I saw that John and Sarah did a lot of head nodding, so I'll give you another chance after Dena.

DENA MENDELSOHN: All right, thank you. Yeah, I just wanted to thank Quinn and Ken for bringing that up. In the immediate short term, we are not getting any privacy laws passed in the next short term, couple months, and so individuals do need to be very savvy in the marketplace. But like everyone else is saying, it does seem quite inappropriate to shift the burden to consumers to do a lot of homework, and it really makes an assumption that consumers are in a position to always protect themselves, when really that is not the case.

Another concern that I also have is that when we tell people to rely fully on privacy policies, we're basically putting developers and manufacturers in the position of creating their own laws and then following them. And then we're expecting of the FTC to be able to enforce on every individual law, which also does not seem reasonable at this point. So looking forward, what we definitely need is for lawmakers to promulgate comprehensive data protection for individuals.

ELLEN CONNELLY: Thank you. John and Sarah?

JOHN TOROUS: We agree. Even from the study we presented, where we showed that the apps aren't really even following their own privacy policies, I wonder if, as laws and legislation eventually take it back, there needs to almost be a focus on educating people to be aware of it, too. I think there may not be the demand for it because I think all of tuned in and listening are aware of these issues.

But I think a lot of times the shock, when you show someone what data an app is taking, again, a clinician, a patient, it doesn't matter who, people actually don't expect that this much has happened, or this type of data movement is happening. And again, I think it's because they say, well, when I'm in a clinic visit, I expect kind of privacy. This app is talking about clinical things.

So I think raising even just awareness among people and educating them is probably a good first step. It's not comprehensive, but there aren't that many systematic efforts to do this. Or even clinicians don't have great resources to turn to learn about these issues. I think, again, it would be almost nice if we could force everyone to watch what happened today. It would probably make a good first step in this.

ELLEN CONNELLY: Thank you so much. I'd like to now change gears a bit and I'm going to throw it to Elisa, who's going to ask a question about the Cures Act. Elisa, I think you're on mute.

ELISA JILLSON: Hi, can you all hear me now?

ELLEN CONNELLY: Yes.

ELISA JILLSON: Yes. OK, great. So as Ken mentioned, following passage of the 21st Century Cures Act, the Department of Health and Human Services issued new rules intended to support patients access to their electronic health information. Some observers believe that these new rules will significantly increase consumers' adoption of health apps, use of health apps, that are not covered by the HIPAA detailed privacy and security safeguards.

What are the implications of your research for the projected shift in how consumers use help apps? From a privacy perspective, how ready is health app universe for this shift? And I guess my last question-- I know many of you have touched on policy implications and where more regulation or different regulation may be needed. But coming back to the research, where is more research needed so that we are in a position to prod the app universe into the right direction?

KENNETH MANDL: This is a fantastic question. From a utilitarian perspective, I have good news. The uptake of the apps economy innovation marketplace has been relatively low so far. That's for a couple of reasons. One is that the regulation is new and doesn't take full force until 2022.

The other is that it's complicated to create these apps and to educate consumers that they even exist and to get them to use them. So from a technologist point of view, that's a big headache. From a privacy point of view, it gives us the advantage in that not that many people are being exposed to this risk yet.

The other aspect of the good news is that, by far, the most common consumer app that connects to this API is the Apple Health app. And to date, Apple, for its health app, has taken an extremely rigorous, privacy-first perspective. Apple does not mine the data.

There is tremendous value in those data that are in those patients' health apps, and Apple leaves it encrypted, available only on the patients' device, and backed up, also encrypted, to the patients', or consumers', iCloud account. So it doesn't look across them. It leaves it with the consumer. And it has a special process, much more rigorous, than its process for general apps, for apps that will access the health data that has been downloaded to the patient's phone.

So the good news, again, is that uptake is slow, and where there is uptake, right now we have a lot of safety. But the issues and the caveats that we have seen throughout these talks are what we are facing not too long from now. And in addition to data that is going to be equally concerning, certainly the data that patients and consumers enter into mental health apps, is no less concerning than anything coming across that API.

Nonetheless, the data coming across those APIs will include, actually, clinical notes and summaries, eventually, hopefully potentially, images, things are very revealing of many aspects of the patient. And I think we need to reinforce what happens as the data traverse those APIs with real standards for privacy policies and real means to enforce them, and tremendous education and research into how patients actually understand those policies, and whether they can follow them and what the real risks are.

The other aspect, I think, is comprehensive privacy legislation so that, on the other end of this, these data that are health-related and health-relevant, are in fact, in some way that the consumers are protected from the use of these data. And that's going to take some real creativity, to come up with legislation that both promotes innovation and also protects patients.

ELISA JILLSON: Thanks, Ken. Do others have anything to add? Are there other areas where additional research is needed to make this app universe ready for us?

JOHN TOROUS: I'll just briefly add, I think we still need to understand what both consumers and patients value in the data, what they are understanding, how people understand what their data is worth, what you're willing to trade, compromise. We're not telling people never share your data, but I think we still haven't, again, educated people on what it is, what they have, why it's valuable, when it matters, more than less. I think, as Dr. Mandl says, the stakes get higher. It's on us to make sure at least everyone is aware. We don't have to put the burden on them, but certainly they need to know what they have.

ELLEN CONNELLY: That Mute button. OK. I think we'll move on to another topic, and I'd like to make some linkages between at least one slide that John and Sarah had up at this conference and some research that was presented at PrivacyCon 2019.

So some observers of the app market have argued that you get what you pay for. Free app sells your data to turn a profit. The paid apps are a bit more privacy protective. Research that was

presented at PrivacyCon 2019 challenged that idea, that paid apps are necessarily more privacy protective than their free counterparts.

And so as I mentioned, John and Sarah, you had a slide on this that suggested similar results from your analysis. I'd like to get some thoughts from all of the panelists about how, does the free versus paid distinction play out in the health app context? And also your thoughts on whether additional research is needed here, and if so, what kind of research. I'd like to start with Quinn for this, and then maybe move on to John, Sarah, and the others.

QUINN GRUNDY: Sure. So I think, yeah, the work that John and Sarah and others have done obviously debunks the assumption that if you've paid for an app, your data will necessarily be private. I think one area that our research highlighted that maybe needs some more attention is the relationship between developers and third parties.

In particular, there are a number of third party services that are used to monetize apps or to enhance the features of an app, whether that's user analytics or area testing or social media integration that are offered to developers in a freemium model. So developers can access these services without cost, and often, that's in exchange for access to de-identify or aggregate user data.

Often, developers who pay for higher tiers of service, sometimes there are different data sharing agreements. The problem is is that consumers have no way of knowing what kind of agreement developers have with third parties, what kind of data sharing protections are in place. And the relationship between the user and the third party is far from transparent, and they actually, in many cases, have no relationship at all. And so I think greater scrutiny and transparency with these behind the scenes relationships needs to occur so that consumers can understand what is ultimately happening with their data, whether not it has their name attached.

ELLEN CONNELLY: Thank you, Quinn, John, Sarah?

JOHN TOROUS: What Dr. Grundy said is exactly correct. I think the business model of apps is a different topic for a different day. But a lot of these apps are moving towards subscription models, so it actually also becomes complex. So they'll have a free version that's a limited trial or limited features, and then you pay to continue using it. So business models are evolving. And there aren't actually that many truly free apps, and the ones that are free are usually just information resources that don't really do much, not in a good or bad way.

But it's also interesting. As the business model of these apps evolve, how does the privacy around them? And when you pay for a subscription, what do you get or not get? I think that's a topic and you have to learn a lot more about, as well as if the employer is paying for the benefit. There's a huge move, at least in mental health, to try to say the employer will pay for this. What does the employer have access to or not? So many open questions.

ELLEN CONNELLY: Thank you. Dena, I'll give you the last word on this topic.

DENA MENDELSOHN: OK. Well, I'll keep it brief, but I just wanted to push back on the notion that a paid app should have better privacy protections than unpaid ones. This could create a major issue, where lower income individuals are put in a position of picking between a free app that may not be as privacy protective versus having to pay in order to get access to, perhaps, a central service, like a mental health app. And so this is yet another reason why we need comprehensive data rights set in law so that we have a baseline that everybody, regardless of income or ability to pay, can expect from their health apps.

ELISA JILLSON: Thank you all for those thoughts. We have just a few moments left, so I'd like to ask if you all have any wrap-up thoughts. We had an audience question about what legislation is needed in this area. I think that's probably a question that would take more than one minute of wrap-up. But if you could briefly, in your closing remarks, address where you think research should be headed and, if you'd like to, where you think regulation or legislation should be headed as well. And we can start-- Ken, why don't you start us off?

KENNETH MANDL: Well, I think I would focus some of the research on this transition across the API from a HIPAA-covered entity under consumer direction to a third party app. There we have a controlled environment and a regulatable environment. And getting that piece right will help consumers enormously in protecting their privacy and their integrity in the face of using apps, and also in helping to prevent misuses of their data.

The research needs to be done in what patients expect at that moment, what they can understand, how much external protection they need, and where regulation versus community standards becomes the most effective focus. But I'll emphasize that because the FTC could potentially be overseeing the regulation of a very large amount of health data for the first time. Data that HHS is used to regulating and the FTC is not yet used to regulating, I think we have an opportunity to really think this through together, as a community and as a nation, on how to make the FTC most effective in taking on this new role.

ELISA JILLSON: Quinn, maybe we can go to you next.

QUINN GRUNDY: I think at the moment, our existing legislation regulation and market place puts the most responsibility on the groups with the least power to do something about this, so consumers, and to an extent, app developers. And I think the focus of regulation or legislation needs to shift to some of these really big players with much more power, including app stores and distributors, data aggregators, and digital advertisers, who currently are very much behind the scenes and engaged in a lot of these sometimes dangerous and harmful practices, but aren't really the topic of discussion at the moment.

ELISA JILLSON: Dena?

DENA MENDELSOHN: I think at the end of the day, it's on our lawmakers to enact legislation that sets a data right framework that could serve as a baseline for health apps and other connected technology. And that way, health app developers can focus on creating the best technology that can win in the marketplace, and consumers could trust that the technology that they've chosen to further their health and their lives will not be used against them.

ELISA JILLSON: Thanks. And John and Sarah?

JOHN TOROUS: It's hard to follow all of that up. So I think we would say perhaps we do need to start using and investing these frameworks in real world setting, and actually, again, educating people, giving them resources they can use today. On a more flippant note, if anyone has a name for the database that we've built, we'd love your help in naming it. Calling it the App Database is a little bit boring. So please send us any names you have. We're open to it.

ELLEN CONNELLY: OK. And with that, we are over time. So I want to thank-- Elisa and I really want to thank all of our panelists for this really interesting discussion and great presentations. We appreciate it. We'll have a short break, and our next panel, which is Bias in AI Algorithms will start at 10:50. Thank you all so much.

DENA MENDELSOHN: Thank you.

BEN ROSSEN: Good morning, everyone. My name is Ben Rossen, and I'm an attorney in the Division of Privacy and Identity Protection at the Federal Trade Commission. And it's my pleasure to welcome you all to our second panel of the day. Today's PrivacyCon is primarily focused on the privacy of health information and mobile apps, but this panel has a little bit of a broader focus on what is a very important issue, and surprisingly is one that we haven't covered in a previous PrivacyCon. Namely, that is algorithmic bias and the risks of data discrimination.

So we are extremely lucky today to have two really terrific panelists. First up we're going to have Mohammad Ali. He is a PhD candidate at Northeastern University, and he's going to be presenting his paper entitled "Discrimination Through Optimization-- How Facebook's Ad Delivery Can Lead to Biased Outcomes."

Next, we are very lucky to have Professor Ziad Obermeyer from UC Berkeley's School of Public Health, and he's going to be presenting his influential paper about bias and managed health care algorithms, entitled "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations." And you could find their full bios on the event website.

We're going to have two 12- to 15-minute presentations, after which there will be an opportunity for some Q&A. And with no further ado, I'm going to turn it over to our first panelist. So Ali, I'll let you take it from here.

MUHAMMAD ALI: Thank you. And my slide's online right now? OK. Well, thank you so much for the introduction, and thanks to everyone who is watching this. So today, I wanted to talk a little bit about discrimination in online advertising. And if you've been following the news, you've probably read an article or two about it.

But a lot of the focus in the past was focused on the targeting side of things, how these online platforms are built in a way where they provide this breadth of options to advertisers, and in essence, enabling them to exclude certain users from seeing their ads. But I'm not going to talk about that.

What I wanted to focus on was the delivery side of things, where once an ad starts running, the algorithm is making decisions on who to show the ad to. So that will be the focus of this talk. But before I talk about my results, I wanted to give a brief climate on what the Facebook advertising system looks like. That's what we focused on in this study. Next slide, please.

So here you can see-- if you have a Facebook account, you can go to the Create Ad option in the top right, and in a couple of clicks, you'll end up on this section. You see you can target by location. There's a bunch of demographic variables here, age, gender, language.

And at the bottom there, you can see there's detailed targeting. These are interests that Facebook is constantly inferring about its users, whether you're interested in coffee or comics, and then they present all of these attributes to advertisers to target. And that has been the focus of a lot of the prior work. Next slide.

For example, these are some of the examples. On the top here, you can see, back in 2016, Republica showed that they could target people looking for housing and exclude people by their ethnic affinity, as Facebook was indexing at the time. And later, where it showed that even if Facebook goes ahead and blocks these features from being excluded, as they later did, a malicious advertiser can go ahead and find other proxies that correlate with race, and then go ahead and exclude that. So there's a lot that a malicious advertiser can do here, but that's not the focus here. Next.

So we look at the advertising system in these two tables. There's the advertiser, who is controlling the targeting part, where they design the target audience, what the ad looks like, how much money they want to pay. But then once the ad is created, it goes to review. The advertising platform is making decisions on which user they want to show these ads to.

And they're running an auction. They're doing some estimates of relevance. We want to understand whether the differences-- any sort of discrimination can arise in this space. So can there be delivery skews on the second phase? Next.

And we do that simply by actually buying ads from Facebook, because there's no clear way-- there's no data set where you have information about targeting, and then the eventual information about delivery. What we had to do, we had to create our own ads, sign up as an advertiser on Facebook, and then ask them how those ads are doing. Facebook is happy to report breakdowns by age, gender, location, multiple other things.

So we use the APIs to collect all this information on the ads that we run ourselves. We thought this was the best way to do this. Next slide. And one of the first set of ads [? we downloaded, ?] these two extremely stereotypical ads that we expected was skewed a certain way. So one is advertising bodybuilding.

The other one is advertising a makeup kit, pointing to Elle or bodybuilding.com, both websites that we don't own. And we target these two ads to the exact same set of random phone numbers in the US to see, given that the targeting is the same, how does the delivery affect? Next slide.

And we see that there's these large differences, where one ad has, eventually, 85% of male audience and the other just has 5%. So it's clear that just the targeting, just the delivery phase, can cause these large differences, regardless of the targeting. Next slide.

So that's the first question that I asked on how these differences can [INAUDIBLE] in the delivery phase, yes. But we want to understand it better. How do these differences even get there? What elements of the ad is Facebook looking at? Are these differences because users are clicking on these ads more? Does this decide a priority? I'm going to try to go through all of these one by one and see. Next.

So this is what a standard on Facebook would look like when you are advertising a link. You can see there's so many things you change here. There's the text on top. There's the image. There's the URL. These are just the user facing attributes. And behind the scenes, there's other activities as well, such as the daily budget, what audience you're selecting. And we wanted to tweak each of these to see what causes the most difference. Next slide.

And we realized that even before we changed any of the interfacing attributes, as I mentioned, just changing the budget itself causes differences in how many women see the ad. So we ran this ad for Indeed, the job search website site, from one of our pages. And we noticed that the more money we were paying, the higher fraction of women in the eventual audience we were reaching, arguably because women are more competitive on Facebook or because they're more expensive, for some reason. But these are differences that the advertiser would not be able to realize what's happening because-- so we stuck to a $20 budget for all of our experiments, so these baseline effects disappear. Next slide.

And then we started to tweak the attributes of the ads themselves. So when we started running this, my expectation was they're running some sort of natural language processing and they look at the text that I put in the ad and that's how they decided who the ad is relevant to. Turns out I was wrong.

We ran an ad with just a baseline, a white image, with a text on the white image. And we see that there's no differences between the bodybuilding and the cosmetic site. Adding the headline causes some differences, but not the sort I mentioned earlier. Next slide.

But adding the image immediately causes these large differences. We see that as soon as we add the image of the guy pumping iron and the bodybuilding, the initial skews that we saw just immediately replicate. So it seems like in these ads we were running, the image is the strongest factor to the classification algorithm and it's relevant [INAUDIBLE]. Next slide.

And one of those other things-- which this was also in our initial hypothesis-- it might be because people are clicking on these ads more, because people are interacting. We pulled the API over the 24 hours multiple times, but it turns out that some sort of relevant estimate was made as soon as the ad started running and the platform sticks to a decision throughout the course of the ad. So there is clearly some initial decision being made. Next slide.

And this was one of the harder things to measure, but we wanted to be really be sure how much of this difference was because of any humans in the loop versus algorithms. By humans in the loop, I also mean users who might be giving telemetry data to Facebook, basically scrolling over my bodybuilding ads differently than cosmetics, or any sort of modulators that might be [? in the loop. ?]

So we wanted to create ads that would make no sense to people but would make sense to an image computer vision algorithm. How we do that is we take images and we try to make them transparent. This is an example of that. You can see that this image looks slightly transparent.

You can see on the right there are RGB values for multiple pixels here. So each pixel has an RGB value, and then the alpha channel, which controls the transparency. And this is slightly transparent because I've turned down the alpha channel all the way to somewhere in the middle. Next slide.

And if I was to turn the alpha channel all the way down close to 0, it would look basically a blank white square to a person. But the computer vision algorithm can take these RGB values and work with them. And it's funny, because I when I sent these slides to the organizers, they were confused. They said, something is missing in these slides. It's built a reverse CAPTCHA, where it doesn't make sense to a person, but it makes sense to a computer. Next.

And we use this technique to basically take images where we knew, working with the algorithm, they were skewed towards men and images that we knew were skewed towards women. And on both visible and invisible images, so that any sort of user interaction has gone away. It's just the image algorithm. Next slide.

And you can see here, for example, the blue colored dots on the top. You can see the hollow ones are the ones where the male images were made invisible. Between the visible and invisible, there's barely any statistically significant difference. So the gender estimate, the gender skew remains the same, regardless of what it is. Because the user is seeing just a plain white square. It's not any sort of data that was being incorporated there. It's just that the image algorithm sees a certain image, it classifies, it, and it sticks to its judgment. Next.

So to-- we'll go through all of these [INAUDIBLE] to gain a better sense of how the algorithm is working. So we understand that it's mostly the image that's causing all these differences. A lot of these differences are made as soon as the ad starts running, and humans are not as [INAUDIBLE] as we thought. And we say at least because we're not sure, because these ads aren't run for weeks or months. So we don't know what would happen if we got hundreds of clicks on them. But at least in the few days that we run these ads, we see that a lot of these decisions are algorithmic. Next slide.

But one of the other things we really wanted to measure was whether Facebook is capable of producing any sort of racial skews and Facebook wouldn't report us breakdowns as it does with the gender, where we can ask the APA for information. So to get at racial information, what we do is we take water records on North Carolina.

So we build this methodology where we divide the state of North Carolina into regions, where we only take information of black waters from the water records and upload that to Facebook to create an audience, and regions where we only take information about white [INAUDIBLE]. So from the water records, we can get information like first name, last name, zip code, and a lot of other things, and we can target these people.

So when Facebook reports the location back to us, we know that we only uploaded black users in this area, so we can infer this. And to test whether this works or not, we run yet another set of stereotypical ads. Next slide.

Where we essentially take the top 30 country albums, top 30 hip hop albums, all pointing to rollingstone.com, the same website, just different articles with images. And we see very, very strong skews, where the country music ad goes to 80% white users in the audience and the hip hop ad is only 12% white users and the rest of the audience is black. So this gives us confidence that this reverse influence methodology that we come up with for measuring this works, and we can use this to measure these facts in more important categories. Next slide.

And by what I mean by more important categories are protected categories, employment, where it's illegal to discriminate. So a lot of the examples that I shows so far, they might be benign. Judging whether someone like sneakers or not doesn't seem too problematic, but doing the same thing excluding someone from an employment opportunity would create some sort of liability.

So what we do is we create these job ads on multiple ads. For example, this is a job in the lumber industry, the cleaning industry. All of these ads point to indeed.com job searches. So if someone clicks on it, they actually go to an actual job search. And we target the exact same set of people for all of these ads. Next slide.

And we see the same differences exist, even for these jobs ads that we saw earlier. For example, on the left, you can see the gender distribution. You can see that the number of job ads are close to 90% male, while the janitor ones are skewed towards women. And on the right, you can see the racial split, and you can see that the lumber jobs skew towards white people and the janitor actually skews slightly towards black users. Without the advertiser ever asking anyone to do so, this is the exact same set of people that both of these ads are targeting. Next slide.

And we see not just in these two categories. We've done it for a variety of jobs-- supermarket workers, secretaries, nurses. And you can see that across gender and race, there's so many differences that occur on the delivery side of things, even when the advertiser might not have intended to discriminate in any way. Next slide.

So in essence, to summarize, what we do is we provide these new methodologies to be able to measure Facebook's advertising system. And we show that, regardless of how an advertiser decides to target, a lot of these differences can arise in the delivery phase. And not just in benign categories. It can also bring into protected categories, like employment.

So what are the real world implications for all of this? And I'd like to mention, last year, the Housing and Urban Development Department, they decided to sue Facebook because Facebook

was enabling discrimination in housing opportunities. So our paper, we believe, provides a way to investigate whether these differences-- how much of the differences arise from the delivery part versus how much of these differences are responsible by the algorithm itself, who's deciding who to show the ads to. So it's a methodology towards that.

We also think our paper sort of provides a unique nuance on the Communications Decency Act, section 230. So this provides a lot of immunity to online publishers from all the content that they're hosting. So it's the responsibility of the people posting and not the publishers'.

But what we show is that if so many of these decisions on which user eventually ends up seeing something are contingent on the delivery algorithms, on the AI that's running in these systems, then it's not so clear who's entirely responsible. And finally, I'd like to emphasize that we're still at the phase where we need more transparency into these systems.

Whenever something goes wrong, online advertisers cannot continue to blame the advertisers for being discriminatory, when we clearly show that so many of these differences don't even depend on the advertising. A lot of these decisions are because these algorithms are optimized so heavily for relevance that they might end up skewing the graphs. Next slide.

Yeah, that's all I have for today. I would like to profusely thank my collaborators, [INAUDIBLE] and Alan at Northeastern, Alexander at USC, and Aaron [INAUDIBLE]. And thank you, again, for listening.

BEN ROSSEN: Ali, thanks so much. Next up, we have Professor Ziad Obermeyer. He's going to be presenting his paper, "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations." Ziad, I'll turn it over to you.

ZIAD OBERMEYER: Thanks, Ben, and thank you so much, Ali. I think that the work that Ali just presented was such an ingenious example of the kinds of ways that researchers have tried to essentially study algorithms in the wild. So if you think about all of the things that that research team had to do to understand what exactly Facebook was doing, and in some ways, probably even better than Facebook understands what they're doing themselves. It's this careful process of pinging the system, seeing what happens, reconstructing results.

And all of this stuff is done, essentially, from the outside. Because in a lot of these settings, when we want to study algorithms that are operating at scale in our society, we can't get inside. And we can't get inside for some reasons that are not so great, like the algorithm developers don't really want us to get inside. But also some reasons that are legitimate, that there are trade secrets and things that we legitimately don't want to make public.

And so I wanted to talk through one example from our work, where we had an enormous luxury, relative to most studies of algorithms, which is that because we were working in collaboration with a health system that had actually purchased one of these algorithms, we could see everything about it. We could see all the variables going into it. We knew exactly what the algorithm was doing. And maybe most importantly, for the purposes of making the case that

there was racial bias, we could actually follow up what happened to patients and document the impact on health outcomes.

And so I think that this one example, or at least I hope, can teach us some general lessons about, essentially, how to be good users of algorithms. And that's on the consumer side, but also on the regulatory side, as we try to make sure that bias doesn't get into these algorithms, and if it does, how to hold organizations accountable.

So our example that I'm going to tell you a little bit of background on upfront is about our system's effort to help complex patients. So in general, our health system does a not so great job of helping people with complex health needs. They often end up in the emergency department or in the hospital, if they're on many medications that often conflict.

And so over the past few years, the health system has gotten very interested in trying to intervene early on these patients. And the idea is if you imagine a person with heart failure, a person with diabetes. There's a window of opportunity to help that person early, when problems are still able to be nipped in the bud.

And so what the health system has invested in very heavily is what's called high risk care management programs, to do exactly that. So the idea is that these patients are treated like VIPs. So patients with chronic conditions are given a special phone number to call. There's a special team of nurses who can make home visits. They can arrange for a next day primary care appointment. So it's really they want a low threshold for these patients to call in, reach for help so that this team of trained experts can nip all these problems in the bud.

And the goal is twofold. The goal is, of course, to help patients so that their health problems don't go from small problems to big problems. And the second goal is to save the health system the money that's associated with those problems turning into big problems, people ending up in the hospital.

So as you can imagine, that SWAT team of specially trained nurses and extra primary care slots and home visits, all of that is fairly expensive. And so you can't do this for everyone. You have to choose your patients carefully. And that's where algorithms come into this story.

So it's fundamentally about resource allocation. We have this scarce resource of extra help programs and we want to target those resources to the people who need it most. And if you think about most health systems, they're managing tens, if not hundreds of thousands, of patients. That's not a great job for humans to do. And so a lot of health systems have started investing in algorithms to at least start that screening process for them.

And so if you take the industry estimates seriously, the scale of this is just enormous. So the industry itself estimates that around 150 to 200 million people are screened by this family of algorithms every year. The particular software that we're using is one of the largest in that market, and so that's what we're studying.

And the way these algorithms are generally used is almost as a first step. So there's a primary care population. And the algorithm just runs in the background and generates a score for everyone in that population, and then the health system does something with that score.

So in the particular decision that we're studying, the top few people were just fast-tracked into this high risk care management program, and about the top half, except that top 2%, 3%, those people were shown to their primary care doctor and the primary care doctors were asked, should this person be in this high risk care management program? So a lot of variety in the institutional practices, but ultimately, the algorithm does a screening step, and then that screening is used to decide lots of things about the patient, but in this case, should that patient be enrolled in one of these programs?

So on the next slide, there's a graph. And I'm just going to talk through it slowly because I'll show you a few graphs that look like this and I just want to make sure they're all clear. So on this graph, on the x-axis on the bottom, is the algorithm. So this is what the algorithm thinks about people, and it's arranged from very low risk on the left at 0 to very high risk on the right. And those top few percent, to the right of that vertical dotted line, those are the people they get fast-tracked or autoidentified for this program.

On the y-axis is a measure of health. So this is basically at a given level of what the algorithm thinks about you, how healthy you end up being in the next year. Concretely, it's a count of how many chronic conditions you have that flare up over that year.

The two lines show two groups of patients. The top line, the purple line, is black patients and the bottom line, in gold, is white patients. And as you can see at every point in this distribution, black patients, at the same score as white patients, have a worse health, on average.

And so I think that violates what you could think of as our working definition of bias. So the algorithm is being used to guide a decision. And so two people who have the same algorithm score are treated the same by the algorithm, and thus by the health system who uses the algorithm. So those patients should go on to have similar health needs, irrespective of the color of their skin.

And what we find is that if you just look at that high risk group, where people are fast-tracked into the program, the algorithm, operating on its own, judges that high risk group to be a group of patients that's only 18% black. When we did a very basic analysis to say, what would this look like if the algorithm had no bias based on need, that number would rise to almost half, to 47% black. So this is not a trivial amount of bias. And again, the definition of bias that we're working with is at the same algorithm score, people should have the same needs, and that turns out not to be the case.

So on the next slide, what we're trying to illustrate is where we think this bias got in. As I mentioned, we knew exactly what this algorithm was doing, what it was predicting, how, what variables. And it turns out that if you step back-- this is a very complex question-- who has health needs?

So in most data sets, we don't have a variable called Health Needs. And so what we do instead is we pick a proxy variable that's measured in the data sets that we have access to. And what the algorithm developers did in this case-- which is a very common choice. This is not just about this particular developer. This is a very common strategy-- is we use costs as proxy for health needs.

Now, that's not unreasonable because, in general, when you're sick, you go get care and you generate health care costs. The problem is that even though, on average, that relationship is true, that you generate costs when you need health care, that relationship is very different for black patients and for white patients.

So when you need health care, you're less likely to get it when you're black, and that leads to lower costs. So in this graph, we're showing you on the x-axis, instead of the algorithm, a measure of health. So increasing health needs further to the right. And what you see is that white patients always have more costs on average, no matter where you are in this health distribution. And in our sample, black patients cost a substantial amount less every year at the same level of health.

So on the next slide is our hypothesis of tying this all together. Using proxy measures is inevitable, but some proxy measures are biased, and we think this is a very common mechanism by which bias gets into algorithms. In our example, it was using cost as proxy, for health and not realizing that costs were just lower at a given level of health for black patients.

But you can imagine many other situations like this. We often use arrests or convictions as proxy for criminality, but that is not an unbiased measure of criminality. We use income to measure credit worthiness, and that's going to introduce all of the biases we already know about and differences in income by ability.

So all of these things, because they're subtle questions about correlations with underlying truth with race, they can be subtle, and that's why this wasn't caught. It wasn't caught by the people who developed the algorithm, even though they were very well-intentioned. It wasn't caught by any of the clients that purchased the algorithm, even though these were people who have a deep commitment to fixing disparities and improving population health. And it wasn't caught by the humans who were either using the algorithm or being affected by it. And that, what you can think of as a market failure, is the reason that I think there's an important role here for regulation.

And so the question is how? And so I'll just leave some of this to the discussion, but I'll just say that where anything starts, making sure that the algorithm that you develop or buy isn't biased, regulating and holding organizations accountable. All of this starts by having a very clean definition of what we mean by bias.

So in our case, it was two patients with the same risk score should have the same health needs because this risk score is being allocated, is being used to allocate a health resource, and it shouldn't matter what color their skin is. That definition is the beginning of lots of methods that you can use to test for bias, to query algorithms that an organization is thinking about buying, and for regulators, to offer guidelines to industry.

And critically, none of these things require compromising trade secrets. All of these things can be done from the outside. We don't need to understand or interpret the algorithm. All of these things can be done with our basic level of data access that we have.

So to wrap up, I'll just tell you that after seeing this work, we actually reached out to the company that developed this algorithm and we worked with them-- they were incredibly responsive and positive-- to replicate our results and their data and to patch their software by predicting a measure that was closer to health and not so close to cost.

And when we did that, we saw really large reductions in bias. And I wanted to mention that because we've expanded this effort out to work with a number of different health systems, insurers, algorithm developers. And our email address is on that last slide, if you want to reach out. Thanks so much.

BEN ROSSEN: Great. Thank you so much. So we have an opportunity for some Q&A. And I know if the folks who are watching on the Livestream have questions, you can submit those by Twitter or otherwise, and we have somebody who's going to pass those along to the moderators.

Ziad, I'll start with you to just get that conversation started, since you hinted at this a little bit already in your talk. What is the takeaway for developers and health care systems and regulators, in terms of applying the lessons from your work as a practical matter? And are some of these applications already out there in the field, given the work that you're doing with health providers and developers in light of your paper?

ZIAD OBERMEYER: Yeah. Thanks for asking. I'll say, first off, that it's such a treat, as an academic, that anyone in the real world is interested in your work, and so it's been a real privilege for us to work with people who are actually doing things in the world to try to understand and solve these problems.

I think it all really starts with coming up with a working empirical definition of what bias looks like. And I think that a lot of the ways that we tend to do this in practice so far are we look at, is there a race-based adjustment? That doesn't guarantee that there's bias. The absence of a race-based adjustment does not guarantee that there's no bias.

So I think really delving into the substance of what the algorithm is doing, what it's being used to do, and then coming up with a context-dependent definition of bias there that we can test empirically is the first step. And so when we're working with these organizations, the first thing we do is we go really in-depth to understand, OK, here's what the algorithm is being used for. Here's the real thing that we're trying to get at. Here's what the algorithm actually does, and is there a difference there?

So setting up a very clean definition of what bias is is the basis for software developers to audit their own products before they go into the field. If you are purchasing an algorithm, you can set up queries to actually answer those questions. If you are a regulator, you can set up a definition for a given application, and then you can hold people accountable to it.

So I think that's really the core of what we did, and I think the work that Ali presented as well. It's really trying to translate the somewhat abstract notion of what bias means into an empirical data-driven definition in a particular data set. And that's hard because there is no automated process that you can do for that. You actually need to really understand how the algorithm is being used and what disparate treatment or disparate impact would look like in this particular situation, and then set up a set of empirical tests following that.

BEN ROSSEN: Thank you. Really interesting. And to follow up from that, given that there is no off the shelf way of doing this, with these types of algorithms that are purchased from third party developers-- which is still, I think, the most common way that a lot of companies are getting their AI tools-- is there a market failure there, in terms of who has the incentives or the obligations to really examine these types of algorithms? And to both the resources looking at, is it somewhere where regulation needs to set in? Or are there steps that your ordinary companies are able to take to evaluate these risks?

ZIAD OBERMEYER: Yeah. I mean, empirically, at least in the case that we studied, and I think in many others, there was a market failure because there was this problem that wasn't caught by anyone. I think the first part of fixing that is actually to put a name on it and to make it transparent that this is a problem.

In all of my conversations with industry, I don't think there's a single software developer who wants to put out a biased algorithm. And so a lot of them are already taking steps to do that internally, but I think because all of us are just learning about what bias looks like in different contexts and what it means, I don't think that there's a consensus definition on how you even do that if you you're the one that's developing the algorithm or if you're the one that's purchasing it.

And so I do think that this is an area where regulatory guidance would be incredibly valuable, because now that there's a lot of attention, there's a spotlight on these issues, nobody wants to be the company that is putting out an algorithm that someone later audits and finds to be biased. So I think having regulators just set out a definition of what this looks like would be incredibly valuable, because as in most things in medicine, prevention is much better than treatment. It's a lot easier and it saves a lot of pain on lots of different sides. And so I think having consensus around what that looks like would be really, really important.

BEN ROSSEN: I realized that I misspoke earlier when I said folks could submit questions on Twitter. The right way to do it is actually through email, privacycon@ftc.gov. So if you have questions, feel free to send an email that way and we'll pass them along here.

Ali, I'm going to turn to you as well. I thought your paper was really fascinating. One of the things that it reminded me of is a story from a few years ago about when Amazon tried to build a recruiting tool. There's only so many stories that are out there about algorithmic biases, but many of the reasons you mentioned, about not having that kind of window into how these things operate.

But they had caught this algorithm that was going to be used as a recruiting tool because they identified, before it was rolled out, that it was systematically discriminating against women,

despite the intentions of the developers and despite every effort they made to try to fix that problem.

And the findings in your paper reminded me of this, because some of these issues, certainly with Facebook, have been identified in the past. As you mentioned, Housing and Urban Development brought a lawsuit, and folks have been looking at this issue with a pretty keen focus on platforms like Facebook. Is there some reason to think that, regardless of their efforts, skewed ad delivery is just an inherent part of using these types of tools? And if so, what should the platforms be doing to respond, or is there a need for regulators to step in?

MUHAMMAD ALI: Yeah. I mean, yeah, that's a loaded question. But I think the case of-- I like that you mentioned the Amazon case. The Facebook case is slightly different because their advertising tool is a one size fits all thing. It's the same tool that's used for political ads, controlling democracy. It's the same tools used for selling sneakers, the same tool that's used for-- so what works in one context doesn't really work so perfectly in another.

But for the Amazon case, it was very easy for them to test it in that very controlled case and see that. But I think now, Facebook has also started to make other tools for the housing and employment ads, where they're trying to actively address this because so many people have brought forth these concerns.

But these differences arise essentially, as Ziad pointed out, because there's always proxies for making up bad metrics. Because they're trying to optimize so heavily for relevance, what works in one context ends up hurting people in the other context. So I think the only way to go forward is to be cognizant that these algorithms actually have an effect on people and then measuring them.

You can only try to iterate on the measurement and trying to fix these things, and realizing that the way these algorithms are designed-- because you're so heavily optimizing for some sort of machine learning metric of loss, or trying to accurately optimize some exact thing, it just ends up picking up more [? layers ?] and hurting people in the process.

So yeah, I think the only way to do that is to iterate on trying to fix it, and I think Facebook is only now starting to get into understanding that, OK, these things actually have harm. So they're now in that phase where the developers are actually trying to measure and counter these things.

ZIAD OBERMEYER: I'll add one just interesting thing about the Amazon example that you brought up, Ben, which echoes Ali's point, is that in some ways, algorithms can actually serve as a very valuable role of exposing bias in humans. So what was the algorithm in that case doing? Well, it was predicting some variant of, is this person going to be invited back to be interviewed by us?

Now, as Ali mentioned, that's a proxy for the quality of the applicant. But when the algorithm spit out these predictions that were predominantly white and male, that actually was like holding up a mirror to the recruitment process, that that was the source of bias to begin with.

So in a funny way, algorithms can actually work to expose these biases in the human processes that are used to train them, and I think that that's a kind of underrated contribution of algorithm. Everyone gets mad at the algorithm, but it's not the algorithm. It's us. It's just reflecting back what we're doing.

BEN ROSSEN: That actually leads me to a question that we received from the audience, which is for you, Ziad, which was about, what were the alternative proxies that you ended up looking at in your work, as opposed to costs? How did you choose them? And is that process of choosing unbiased proxies something that is replicable?

ZIAD OBERMEYER: Yeah. It's a great question, and I think that it does go back to understanding exactly what we want the algorithm to be doing. So we want the algorithm to identify people in whom we can intervene early and make a difference. So from that point of view, it's actually not obvious that you want to be predicting total costs.

Total cost brings together a bunch of things that you can think of as like good costs, like people taking insulin, which costs money, and bad costs, which are things like people getting their toe cut off because they didn't take their insulin, which also costs money. So when you put those together into a total cost metric, you're conflating a bunch of things that are not the same.

And so what we did is we came up with a metric of avoidable costs, so things like getting your toe cut off because you didn't take your insulin and not the insulin itself. We also have lots of different measures of health that are applicable to different populations, and some are not. So it took a lot more work, just substance knowledge-intensive work to come up with these.

But I do think that in most of the data sets we use, there's a rich set of alternatives. Some are more work than others, but I think the message from our work is that that extra effort can be hugely valuable because it can make the difference between a biased algorithm and one that actually works against the structural biases in our society.

MUHAMMAD ALI: I'd like to go back and talk about that, because it's very interesting what Ziad said about how these algorithms sort of hold a mirror to us and tell us how we're being biased. I really like that argument, but I hate when computer scientists use that argument to just evade all sort of responsibility.

I think a very common thing that computer scientists do is that, oh, the algorithm isn't biased, its it's the data that's biased. But I think it's that very point where-- someone who's trained as a computer scientist, who's been in way too many machine learning classes, it's important to understand that just because the data is biased doesn't mean you let the thing go through.

It's that very opportunity where the algorithm's holding a mirror to you to understand that you're now automating this harm that was accumulated over years. And that's where you need to start auditing these systems. As Ziad said, you need to have clean definitions of bias and work with those until you reduce that harm.

ZIAD OBERMEYER: I think that's a great point. I think there is a tendency to throw up our hands and say, well, we can't have algorithms because the data are biased and the data are biased because our society is biased. And all of that is true. But with a lot of work to take into account structural biases and historical inequalities, we can actually make the difference between good algorithms and bad algorithms.

BEN ROSSEN: I know we're running out of time, and I think that's a great place to end the conversation. I know our next panel picks up immediately after this one. I want to give a big thank you to both of our panelists. Really fascinating work. The papers are available on the ftc.gov website. And thank you so much for having us.

ZIAD OBERMEYER: Thank you.

PHOEBE ROUGE: Hi. So this is Phoebe Rouge, and today, for our third panel, we're going to be talking about privacy and the internet of things. We have three presenters here. The first one we're going to have, Daniel, who did his research at Northeastern University, and he's going to be talking about his research to look at the network traffic from various internet of things devices.

DANIEL DUBOIS: Thank you for the introduction. Yeah. So now I will talk about information exposure from consumer devices. And I will also thank my collaborator, Northeastern University, Jingjing Ren and David Choffnes, and Amy [INAUDIBLE], and Amaria [INAUDIBLE] and [INAUDIBLE]. Next slide.

Usually I start presentation by asking the audience if they have any IoT device. Typically, the majority says no, but actually, most of them realize that they bought a TV in the last 10 years. That TV is likely a smart TV and that can be connected to the internet and that's a full IoT device.

So what motivates this work is that IoT devices have access to private information. They have the sensors. For example, smart speakers can listen to you, like smart camera, smart doorbells, and watch you because they have a camera. And smart TV knows what you do, for example, what TV programs you watch. So all this information is actually shared with their own companies, those IoT devices. And their main purpose is actually to be internet connected, so there is a potential of privacy exposure from that.

And we have seen that that actually happens. The press has actually wrote many articles where there are devices, for example, sharing audio with Amazon workers and other persons like that. And this problem is very important because there are around 10 billion IoT devices deployed currently. And we want really to understand what they are doing. Next slide, please.

In this work, we focused on the devices that are typically deployed in a smart home. We call them smart home devices, like appliances, smart lights, and other devices like that. So what we are interested in is to understand what the device are doing.

What does this mean? We want to understand, what is the destination of the traffic of them? Are these devices talking to their current companies or are they talking to some other third party? And also, is the traffic staying in the country where it is generated, or is crossing the geographical boundaries?

That's important because having state regulations may be different in another country, and sometimes even within one country, it may be different if the traffic stays inside or travel. And also, we want to understand, is the traffic protected by encryption or not? What information is being sent? This is important because if a company is sending private information, it's likely that the user maybe is not aware of that.

So we want to understand where things like that are happening, and also, if any information has been sent unexpectedly. For example, if you have a smart speaker, most people know that they have a microphone, but that microphone should not be transferring the voice all the time, but only when it's used. So we want to understand if that's true or not. Next slide.

Answering those questions is not easy. It's actually a hard problem to measure privacy from IoT devices. And the reason is that the devices are typically black boxes that are much harder to analyze than mobile apps, for example. And the reason is that manufacturers don't provide specifications, and probably, for intellectual property reasons, all the information of how they work is not disclosed.

To overcome this problem, we want to use some technology. For example, we want to employ this initial analysis and information inferences so we know at least what they are doing without having to look inside the device. In addition to the techniques, we need a proper tool to do that, because at the time of this work, there were no tools available to analyze those devices. So to solve this problem, we created some software that is able to collect the traffic and analyze that from IoT devices.

And we deployed this software in two IoT labs, one in the United States and one in the United Kingdom. The one in the United States can be seen on the left, the picture on the slide. And that lab is actually furnished as a studio apartment, where all the devices are put and arranged in a way that are easy to use for their intended purpose.

We actually recruited like 36 students to use that lab. Of course, they signed a consent form. Then we could use their data to see how the device are using for their intended purpose. Next slide, please.

So the device that we consider are home IoT devices, in particular smart cameras, smart house, home automation devices, like [INAUDIBLE], smart thermostats, smart TVs, smart speakers, and many types of appliances from smart freezer to smart vacuum cleaners, for a total of 81 devices.

And we were able to run 34,000 controlled experiments with these devices that were partially automated with our software. We also monitor how those devices behave when not being used. For each device, we monitored 112 hours of [? inactivity ?] to see what they are doing. And

finally, we also looked at what do the devices do when they are actually used? We used a study participant and we monitored them for six months. Next slide, please.

So once we set up our environment and our analysis framework, now we want to answer the question that I said before. And the first line is, where is the IoT network traffic going? We have a lot of plots for a lot of stats, where it's going. But what is important to know about this is that the traffic is actually going to some entities that are not the main manufacturer or the parent company.

Most of the traffic is going to other companies. Most of them are cloud services and [? CBM ?] providers. This is not necessarily a problem, but still, the traffic is still going under control of another entity. What is probably more interesting is that some of this traffic-- like we have seen situations where the traffic is going really to a completely wrong company.

For example, imagine that you have some smart TVs, or at least most of the ones that we analyzed, and it's contacting Netflix. Doesn't look strange, but imagine that you never installed Netflix. You never open it, and you never logged in, and that TV is still contacting them. So that is a bit of a problem, and with a profound [INAUDIBLE] devices under test.

Also, we have seen that the majority of the devices send traffic to another country. 56% of US devices contact other countries and 84% of UK devices contact other countries. Strangely, the UK devices contact a lot of US destinations. So this looks strange, but probably not too much if you think that those devices are typically developed by some smaller companies that maybe don't have the means to create an infrastructure in every region. But still, there are different regulations that apply in each of these regions, and we don't know how this is compliant, like with the US and also European regulations. Next slide, please.

In addition to the destination, we were also interested to the traffic itself. Is this traffic encrypted or not? At the beginning of the presentation, I said that most of the traffic is encrypted, so it's really hard to understand how the devices are behaving. But we analyzed more in detail, and we have seen that a lot of traffic is encrypted. Are a lot of traffic is unknown. That means that we don't know if it's unencrypted or encrypted, but it's still encoded in a way that cannot be read.

If you are optimistic, we can see that it is encrypted as well, but some investigation has to be done. But still, there is some traffic that is red in the figure that is unencrypted, especially from cameras, that are some of the cheapest devices that you can buy. We looked at designing unencrypted traffic, and we've seen some negative trends. Next slide, please.

So that negative trend is that a lot of devices across many categories are actually leaking unique identifiers, like Mac addresses and device IDs. Also, other content is being sent unencrypted, like some actions from simple devices, like turn on and turn off, [? frequent ?] updates of the activity, and also when the device was set up for the first time, which behaves differently from when it's used later. Next slide, please.

In addition to unencrypted traffic, we wanted to see if the encrypted traffic is also carrying some information. And the answer is yes. How did we do this? Well simply, we look at our

experiments. We tried to see how the traffic looks like when a camera is used to produce a video, and then we infer some patterns from this traffic and use these patterns to recognize when the video was sent in our traffic.

And by applying this methodology, we have seen that more than 90% of the devices that we tested that are able to produce a video or voice actually make this information from an encrypted traffic by using our technique. So one problem of this is that this technique can also be applied by any other entity.

For example, internet service provider has access to all the traffic that is produced in a household where the IoT devices are deployed. So they can infer activities and they can see what is done and what is not by those devices, which is a violation of privacy. Next slide, please.

The last question we wanted to ask is to answer if the devices behave unexpectedly or not. We have seen some cases where the device behave unexpectedly. One of them is from popular doorbells, from actually different manufacturers. This doorbell will actually send a recording of the video when a person was moving in front of them.

This feature was not documented at the time and was not even possible to disable. So just owning and using those devices means that the device is self-recording when user don't expect that to happen. We are also seeing cases of smart TVs, not just [INAUDIBLE], but also other companies, that are not related to the apps that have been used during our experiments, such as Google and Facebook.

And last, but not least, we have seen some very popular smart speakers being activated when you actually don't use them. Typically, they have a [INAUDIBLE]. For example, Alexa can activate as my speaker, but unless they activate it if you say something that is different. For example, you could say, I like something, smart speakers activate. So this might just be an limitation of the device, or maybe the manufacturers really want to know what you like. So when you say, I like something, the device activates and sends a recording.

We're seeing other cases of unexpected behavior. For example, like motion sensor, reporting motion when there was no motion, or devices spontaneously stopping or reconnecting. Those are all problems because when the device reconnects, they send all the information again, so they get more chances for violating their use of privacy. Next slide, please.

So all our findings of this study have attracted the attention of the press, so they wrote some articles that actually became very famous and attracted the attention also of the manufacturer. I will say later how we engaged with them to improve their devices. Next slide, please.

So in summary, all the devices that we analyzed had some sort of problems, and the most important is that 57% of the devices, or 56%, have no manufacture destinations, or they send traffic to destinations abroad. This is something that is unexpected.

And also, the vast majority of devices, 89% in case of the US, are vulnerable to activity, for instance, meaning that a profile can actually be created for the users of the devices and how they use them by whoever has access to the network traffic, like the ISP.

This work had some [INAUDIBLE]. As I said before, the press cover some of our findings and the manufacturer contacted us to get more information about why the devices are conducting metrics, for example. We provided them all our information, along with our experiment, so that they could double check.

We never got anything back, like yes, we [INAUDIBLE] still or we don't. But at least they are aware of the problem and will see that some of the latest version of the devices actually have improved a lot, compared to when we performed this study. Also, all the software we produced is publicly available on the website that you see. It can be used to create, for example, new [? testing ?] maps, and we are aware that there is one in Italy that has been built.

And all the software we collected from all the devices can be used to perform further studies, or by the companies to understand how the devices behave. And all this data is also available on the same website, and has already been downloaded more than 100 times. So this concludes my presentation, and feel free to ask questions during the panel [INAUDIBLE]. Thank you.

PHOEBE ROUGE: Sorry. Yes. Thank you very much, Daniel, for your presentation. That's so very interesting and a little-- so there is a lot of information out there, clearly, from the previous talks today, and there's a lot for consumers to understand. So Pardis is now going to talk about her work with Carnegie Mellon and trying to package that information in something like a label so that consumers might these things better.

PARDIS EMAMI-NAEINI: Thank you so much, Phoebe. Hi, everybody, and thank you for joining my talk. I'm Pardis Emami-Naeini, and today I'm going to talk about our project to specify the contents of an IoT privacy and security label. This is a joint project with my colleagues Yuvraj Agarwal, Larry Kramer, and Hanan Hibshi at Carnegie Mellon University. This work has been recently published at IEEE's symposium on security and privacy or S&P 2020. Next.

IoT devices are everywhere. Some of the most common ones, which you might also have at home, are voice assistants, smart doorbells, smart security cameras, smart thermostats, smart toothbrushes, and smart light bulbs. Next.

And some less common ones are smart salt shakers, smart forks, smart umbrellas, and the most controversial of all, the smart toilets. And the list goes on and on. Next.

People are increasingly purchasing smart devices. However, despite the surge in purchasing them, consumers are concerned about the privacy and security of the smart devices they purchase. Next.

And people should be really concerned about these devices. After all, there's been news on how easily security cameras are getting hacked. But sometimes risk could have been mitigated if

users of these devices were more informed. For example, after [INAUDIBLE] security cameras got hacked, the company emailed their millions of users to use multi-factor authentication. So maybe these devices could have not been easily hacked if users knew about better and more secure authentication mechanisms. Next.

You may have also heard about Google putting its consumers at risk by forgetting to mention that its Nest secure hub had a microphone, or in other words, failing to inform consumers about the device sensors. Next.

Another example shows how [? cranky ?] manufacturers are not transparent about their privacy and security practices, is then some smart TVs are selling our data to third parties without disclosing it. Then it got revealed that Amazon is sharing unencrypted recordings of users' voices with its employees. Therefore, in many data collection scenarios, consumers are not informed about who their data is being shared with or sold to. Next.

So what we need here is to find an effective way to show this information to consumers. And this is what we explored in this paper. We designed a privacy and security label for smart devices, somewhat similar to nutrition labels for foods. Our design label covers various privacy and security attributes related to the smart device. And as you can see, we include some of the important information about the IoT devices that IoT companies are not disclosing to consumers, such as access control, sensor type, data sharing, and data selling. Next.

Several pieces of legislation have been proposed, both inside the US and in countries outside of the US, including the UK, Singapore, and Finland that would require IoT labels. So I'm going to mention a few factors that should be included in these labels, but they don't contain too many details about what the labels should look like.

And as you can see from the headlines, these proposals are primarily focused on security attributes without much attention to privacy practices. So our question here was, what should be included on an IoT privacy and security label? Next.

To capture a holistic view, we invited a diverse sample of experts from industry, academia, government, and NGOs. To elicit expert opinion on the privacy and security factors, we followed a three-round Delphi process. In the Delphi method, the objective is to reach a consensus among a panel of experts without those experts directly influencing each other's opinions.

This consensus is usually reached by conducting multiple rounds of interviews and surveys. In Delphi method, we have this concept of controlled feedback, which means that the aggregate output of the previous stage will serve as the input to the next stage. We have this feedback loop to allow experts adapt their responses and eventually converge. Next.

The first stage of the Delphi process is usually an interview study. We conducted semi-structured interviews with experts and asked them to specify the most important privacy and security attributes to include on the label. These interviews resulted in 47 attributes that at least one expert wanted to see on the label.

We then conducted the first follow-up survey. Each expert was randomly assigned to review one third of the attribute and then specify their importance, as well as the reasons supporting their decisions. From this stage, we found the most common reasons for including or exposing a factor.

And then we presented these aggregate reasons to experts on the second follow-up survey. And this is where we have the controlled feedback process. On the second survey, each expert was randomly assigned to review one third of the attributes. And once again, we asked them to specify whether they would like to include or exclude the factor now after looking at all the reasons from the previous stage.

To analyze the interview responses, as well as the opening answers from these surveys, we conducted thematic analysis, which is a recommended qualitative analysis approach, but information is high in subjectivity. We followed a six-step procedure recommended by Plano Clark to create the code book, find the themes, and merge them. Next.

Experts acknowledge the value of the label in informing consumers purchase behavior. An expert said, "What's good about the label is that it empowers the consumer to make a more active decision about cybersecurity, rather than just being completely helpless as to what the security offer device might be."

The average consumer doesn't have a privacy, security, or legal department to review the stuff before they buy. Enterprises do, but consumers do not. So someone's got to be looking out for consumers and giving the consumers this information. Next.

In addition to informing consumers' purchase behavior, some experts reported that the label could be a forcing function for manufacturers to be more accountable and transparent about their privacy and security practices. Moreover, experts mentioned that if the labels get adopted, it could initiate a competition in the market for manufacturers to enhance their practices. And I should mention, there is value in forcing the company to write [INAUDIBLE] down, even if the consumer doesn't understand it. If you said these are your open ports, there would be an incentive to make them few. Next.

As I previously mentioned, experts wanted us to include 47 attributes on the label, which is clearly too many to show on a typical product package. Therefore, we designed a layered label with two layers. The primary layer is the concise format of the label, which could be printed and attached to the package of the product.

And then there is a QR code and a URL at the bottom that directs consumers to the second layer, which has more detailed information and is in an online-only format. Online formats means that it can be updated as the firmware changes, which is critical, as devices get updated often. Another important reason to have this online layer is to have a [INAUDIBLE] to accommodate companies updating their privacy and security practices. Next.

Some of the attributes included under primary layer, [INAUDIBLE] update lifetime, type of collected data, availability of automatic security updates, and availability of default passwords. Next.

Second layer has all the information from the primary layer and a lot more. Some of the attributes presented on the second layer were retention time, data inference, data storage, and whether there is any special data handling practices for children's data. Next.

To assist our labels risk communication and information comprehension, we recruited 15 IoT consumers and conducted a one-hour semi-structured interview with each participant. In this interview, we first asked participants to take a look at the package of a smart device with our label and define the attributes, as well as your values. We also asked them to specify the information that conveys risk to them.

We then asked participants to imagine doing comparison shopping for a smart device from two different companies. We asked participants to compare the labels and specify which company had implemented better privacy and security practices and why. Next. By following a user-centric design process, the [INAUDIBLE] of it improved the design of our labels, and this is the version of our label from last September.

In addition to the label, we prepared a specification document for users and IoT manufacturers. The content of our specification is based on the previous studies we conducted with experts and consumers and several IoT privacy and security references. In the specification, we provided the taxonomy of the label, consumer explanation for each attribute, list off the items to include as additional information for each attribute, and a list of best practices drawn from various references. Next.

The real world impact. We would like to have our labels adopted. And to ease the process of generating labels, we developed a tool that allows users to complete a form for different sections of the label and see the label being generated in real time. In the most current version of the tool, users can download the label in the format of JSON, XML, and HTML. Users can also work on the label offline, and then upload the saved JSON file to resume working on it. Next, please.

To recap, consumers are concerned about the privacy and security of smart devices they purchase. And these devices are not transparent about their privacy and security practices. A label could be useful to provide that much needed transparency, and from consumers purchase behavior.

Although a few proposals advocated for having an IoT privacy and security label, they are not clear about what the label should look like. I showed you some of these legislations in previous slides. To specify the content of the label, we conducted interviews and surveys in a diverse sample of privacy and security experts, and identified 47 pieces of information our experts wanted us to include on the label. To fit this information, we designed a layered label, and what you see on this slide is the most recent version of our label.

To make the content of the label accessible to consumers, we put the most critical information on the primary layer and additional information on the second layer. To ease the process of label adoption and generation, we prepared a specification document, as well as a tool, to generate the label.

And now we're currently looking for manufacturers and retailers to participate in a pilot deployment of the label for their products. So if you want to show your [? agreement ?] to security and privacy, this might be a great start. Please visit iotsecurityprivacy.org to know more about this project and design your first IoT privacy and security label. Thank you.

PHOEBE ROUGE: All right. Thank you very much, Pardis, for that presentation. So next we're going to have Danny, who's assistant professor at New York University's Tandon School of Engineering, present some work on IoT Inspector, which is a tool that collects crowdsourced information on actual IoT, what the IoT devices are transmitting in real time out in the wild.

DANNY YUXING HUANG: Thank you, Phoebe. So hello, everybody. I am Danny Huang. I am assistant professor, starting fall of 2020, at New York University. So as the previous two panelists had talked about, we are constantly surrounded by smart IoT devices, like cameras, Alexas, smart TVs, whatnot. These devices could be constantly watching us or listening to us, but today I'm going to talk about a way for us to watch these devices instead.

So as you see in the next slide, here's a video of me watching Roku TV. On the top corner, on the top half of the screen, is the Roku TV, running the CBS app. I'm just opening the CBS app and watching the live news streaming, without doing anything.

At the bottom is a screenshot of the network activities of the CBS app on Roku TV. I'll talk about how I obtained this screen shot a little bit later. But here's the big takeaway. On the y-axis, vertical axis, is the number of bits sent and received per second. On the x-axis is the time, sped up at 10 times the speed. And each colored bar corresponds to some third party advertising and tracking services that the Roku TV is talking to at the moment.

So remember, here I'm just passively streaming the CBS news, without doing anything on my Roku TV, and the TV is talking to three or four different third party advertising tracking companies. And one of the biggest ones is actually showing in pink. That is actually the Adobe Marketing Cloud.

It's a little creepy, right? I'm not doing anything, watching TV, and my TV is watching me and talking to a bunch of advertising and tracking companies. So in general-- next slide, please.

--there are lots of concerns about IoT security and privacy, not just smart TVs, but Alexa, smart light bulbs, cameras. And as the previous panelists aptly summarized, we don't know what's going on. It's a black box. We don't know what data is being sent. We don't know to whom the data is being sent to. And we don't know even from which IoT devices this data is coming from.

In general, there are two main problems, one for consumers, one for researchers. For consumers, these smart devices are like black boxes. We have no idea what they're going on behind the

scene. And there aren't very many good tools. If you want to start a Wireshark, good luck. It takes some time to set up a wire shark to analyze a network traffic. So that's the first problem for consumers.

The other problem is for researchers. Many research projects on IoT security privacies are limited to lab settings. Like secure researchers would buy a bunch of devices, like maybe dozens of devices in the lab, and connect them to the network, analyze the traffic over a Wireshark, and analyze the traffic.

The problem is that there are more than dozens of devices. There are literally thousands of smart devices in the world, and how to scale the analysis to thousands of different kinds of devices in the world remains an unknown problem.

So to solve these problems faced by consumers and researchers, our vision-- next slide, please.

--is to develop a simple tool for consumers. Our vision is simple. We want to build a piece of software that provides volunteers with usable insights on IoT security privacy with one click. Here, there are two sets of colors, one-click and software.

We wanted to make a tool that is simple to use. No hardware needed. No access point needed to be setup, but something they can download with one click. That's the first vision.

Second vision is usable insight, in green. We want to incentivize users to use our product. It's not just a research project. We want it such that users would want to actively download this software to find out more about their smart home devices, whether my camera is talking to some third party. That's useful insight.

So to provide this vision, we developed a tool called IoT Inspector, which you can download right now at this particular website on the screen. It is Windows-only for now, but we're coming up with Mac in the next version soon. Next slide, please.

And here's what IoT inspector does. At a very high level, it is a tool and it provides a data set. We launched a tool in April 2019. We've gathered more than 55,000 anonymous users at this point, and we're still gaining users and collecting data.

Our users are anonymous, but some users have come out and told us that they're using the IoT Inspector. Examples include reporters from NPR, from Washington Post, New York Times. Some of these reporters are trying to analyze smart devices themselves, but lack the technical expertise, and they use our tool, such as in the case of NPR.

There are other users coming from, say, for instance, Consumer Reports, who told us that they're using this software. And the New York City Cyber Command emailed us and told us that they're using this software to analyze smart devices as well. So it's a tool that's currently being used by thousands of users. You can try to download it, too. Just google for a Princeton IoT Inspector.

So in addition to being a tool, we provide a usable data set for security researchers. In particular, since we launched the software in April 2019, we've collected network traffic data from more than 55,000 internet connected devices. And we've attracted attention for more than 10 research teams requesting data, including academic and non-academic researchers.

Academic researchers include NC State, CMU, University of Illinois, and University of Chicago, looking at different aspects ranging from, say for instance, the privacy of smart devices, like say, for instance, what companies devices are talking to, to security aspects. For instance, I'm working with a group at UChicago, trying to build a smart firewall to protect users from anomalous IoT devices. So essentially, we are doing a service for the community, not just for consumers, but also for researchers as well.

So how does IoT Inspector work? In particular, how do you use IoT Inspector? In the next slide, I'll show you how to download and run IoT Inspector in a test environment. Here, I'm showing a screenshot of Mac. Again, the Mac version will be coming out soon.

So essentially, you download this executable from a website, and instead of double-clicking in the finder window, you right-click and select Open. The whole reason is that this offer is currently not being approved in the Mac app store yet. I'll tell you why it's not approved at the Mac store. You right-click, and then a dialog box pops out asking you if you want to open it. You click Open. And then finally, a browser windows will show a list of IoT devices on their network.

So it's a little bit small here. I'm going to walk you through this particular screenshot of the browser window. So this is the browser window with IoT Inspector's main screen. In particular, it shows a list of devices on my network, like a Wemo smart plug, a [INAUDIBLE] camera, a Amcrest camera, et cetera. For each of these devices, you can inspect in real time what party it is talking to and how many bytes it is sending and receiving and whether the company being talked to is an advertising tracking company.

So again, I'm going to play the same video as I play earlier in the opening slides. Here's a video of me watching Roku TV. Top half screen is Roku TV screenshot and the bottom half is a live screenshot, real-time screenshot of IoT Inspector as I'm streaming the CBS news app on Roku TV. Here the video is a sped up 10 times, but you can basically see the CBS app talking to, basically, four different advertising tracking company, the biggest one being the pink one, the Adobe Marketing Cloud. So beyond just showing the live view of smart devices-- next slide, please.

--we can see, basically, devices that communicate with advertising services. So, in this slide, we see a list of devices and the remote parties that are advertising tracking companies. The right-hand column shows devices. Under My Account, I have a Samsung Smart TV. I have a Google Home. I have an Alexa in my home.

And then I can see on the left-hand corner, left-hand side, the remote parties, they're identified as advertising services. Say, for instance, Samsung was talking to Samsung ACR, a double click.

So basically, the Samsung Smart TV is talking to Google Advertising Services. So the question is how does IoT Inspector work to gather this insight? So next slide has the answer.

At the core, IoT Inspector analyzes IoT network traffic through ARP spoofing. And we'll explain to you at a very high level how ARP spoofing works. So imagine we have a smart camera in the house and it is talking to the internet through our wireless router. Normally, without IoT Inspector, you would have to have your home network router to capture the traffic.

But for us, we make the process simple. IoT Inspector, as it shows in the next slide, captures traffic through ARP spoofing. Here, we have an example of IoT Inspector running on a MacBook. And IoT Inspector was sent to the camera to say, hi, camera. I'm the router. IoT Inspector will also tell the router, hello, router. I am the camera.

In doing so, the computer that runs IoT Inspector convinces the router that it is not a computer, it is a camera. At the same time, IoT Inspector convinces the camera that it is not a camera, that it itself is not a MacBook but a router. So this allows the traffic between a camera and a router to be intercepted by the MacBook that runs IoT Inspector. IoT Inspector can, at this point, see the traffic going through between the camera and a router without actually rearranging cables or setting up a different wireless network. Basically, a one-click solution.

In the next slide, I'm going to show you examples of findings from real devices from real users by IoT Inspector. Basically, there are two areas of findings. One is security, one is privacy. And this is just the tip of the iceberg. I'm just going to explain a few examples.

So security-wise, we found the lack of encryption on many smart devices, including devices made by big manufacturers, like Google and Amazon. Smart TVs on Amazon, some of the apps don't really use encryption. They use, in some cases, just HTTP. And in some cases they use encryption, but they use-- surprise, surprise-- SSL 3.0, which is basically outdated encryption.

We have seen many devices with open unused ports, like cameras that have ports open on port 22. Like SSH, they are never used. But having unused open ports opens up opportunities for exploits by attackers. So these are some examples of security insights. In terms of privacy, we found evidence of advertising and tracking on many smart TVs, including Roku and Amazon.

It also found across device traffic. In particular, your IoT devices did not only talk to the cloud, they talk to each other. So basically, one device can potentially gather private information from another device without your knowledge. So again, more details about these examples in our paper. Just visit our website at iotinspector.org.

So in summary-- this is the last slide-- we built a tool that provides transparency. It is usable, too, for consumers, being used by more than 5,000 users, and we've collected a larger scale data set for researchers to conduct IoT-related research.

But beyond transparency, we want to create action. Next version, where we will build IoT Inspectors such that it will alert users of any actual problems and will protect users from these problems. It is our hope that soon, IoT devices will no longer remain as a black box, but we will

be able to provide transparency and we provide actions to protect consumers. With that, thank you, and I'm happy to take any questions.

PHOEBE ROUGE: Thank you so much, Danny. So yes, if you have any audience questions, you can send them to the privacycon@ftc.gov address and we'll try to get to them. To start out our discussion, though, I will first ask a question of Danny. So one of the questions you might have, as far as using your IoT Inspector, given it's doing this ARP spoofing-- I don't know what that is. So does it potentially introduce any-- would it cause any problems on my network? Would it affect any of the devices in any way?

DANNY YUXING HUANG: Yes. If you want to download on the website, we have a big warning, where it's saying that it's going to slow down the network. Basically, instead of having traffic directly going through your router, as in traffic directly coming from your smart devices to router, it takes an additional hop to your computer, and that's going to slow down your network. That's going to slow down your smart devices.

So if you're, say, watching Netflix on a smart TV, you may experience degradation of traffic. You may still have HD content. You may see some blurriness. This is from real experience of me running IoT Inspector myself in my house.

PHOEBE ROUGE: So this is a question for Daniel. Since you've been looking at this traffic that's coming from these smart devices, have there been any really surprising results? I know, Daniel, you talked about, for example, the smart speaker activating when you weren't expecting it. But have either of you seen anything really notable that was very egregious, or anything like that in your results?

[INTERPOSING VOICES]

DANIEL DUBOIS: Yeah, so we have seen something that was not bad. We have to investigate that. I don't know if I can name the device of our companies now, but there was one doorbell-- not the most famous one, but one that can be bought on Amazon-- that was encrypting the traffic without verifying the certificates. So that means that if you encrypt in that way, encryption is completely pointless. You can actually do man in the middle attacks on the device and get [INAUDIBLE] device and password. So if it happens with a smart camera, that's a problem.

And in general, we analyze a lot of categories of devices. And smart cameras, for this point of view, are one of the ones that behave in the worst way. Because also, the companies are very small. They typically just buy hardware that is made by another company.

They also use software that is by another. They customize it a bit. So it's often updated software full of bugs, and it's concerning. Besides the unencrypted-- in the fully unencrypted traffic [INAUDIBLE] some camera's sending traffic to other residential addresses.

We try to understand why this was happening. We don't know, but they mentioned that your camera is contacting a bunch of addresses. At first, we thought it was hacked. Maybe it was part of the bottleneck, but we didn't find any evidence of that. We might think that the device was

probably uploading some of this data to other device in a way to reduce the use of their computational systems.

But a question about development. We see things that are strange, but it's really hard to see what they do because we don't have control over this [INAUDIBLE]. We just see the traffic that is strange. And the only thing we can say is that when you consider IoT devices, think about what you do for a mobile app. You can install an app from a small company. You don't know what this app is doing. It may send traffic.

And what you can do is if you really need to use the app, just use it, knowing that you are exposing yourself to [INAUDIBLE]. Or you can basically delete the app. That means you unplug the IoT device from the internet. Sometimes they can still be used without being the internet. Also, smart cameras, sometimes they allow you to use them on a network that is isolated from the internet and they still work. So those are only the possible ways that come to my mind where consumers can protect themselves on this equation.

DANNY YUXING HUANG: And some of the surprising things that we found is actually from smart TVs. One example is that, say for instance, the Roku-- I'm sorry, the Amazon smart TV screen, for instance, has a built-in feature that basically says you can actually opt out of interest-based advertising.

If you think that turning this off, turning off the interest-based advertising would reduce tracking, you're wrong. So in one experiment, we found that we turned off interest-based advertising on both Roku and Amazon. We still see these devices potentially sends the information to some third party advertising tracking services. So yeah, it's one example. Tip of the iceberg for some of the privacy issues we found in smart TVs.

PHOEBE ROUGE: OK. So I guess following on, Daniel, you mentioned some things that you might want to do if you get one of these smart devices, to address some of the concerns. I guess I'll start with Pardis. If you are a consumer that wants to buy a smart device, and I'm watching this PrivacyCon and I'm like, wow, there's a lot of things to be concerned about, what's the first thing that you would look for?

Like if I bought a smart device, what's the first thing I should do if I unpack it? Is there any setting I should change? Is there anything I should look at to make sure it does or doesn't do, anything along those lines.

PARDIS EMAMI-NAEINI: That's a very good question. So I think, basically, privacy really depends on your own preferences, definitely. So you may be concerned about some type of data and you may not be concerned about other types of data. But apart from that, I think what is really important for consumers to know about is to know what types of controls they can have, if they want to change them or not.

So basically, when you purchase a smart device, I think the first thing that you should do is to understand the settings of the device, the privacy and security setting of the device, to basically

know how you can change data sharing, how you can opt out from data sharing, data selling, for example. Do you have this option?

And another, I think, important thing is to understand the basics of privacy and security information of the smart device. For example, whether the device is the default password, or whether you would get security updates. So there are some critical information, privacy and security information, some basics that you should really know about. And then other than that, the types of controls that you can have. So I think that's the first things that I would recommend consumers to do.

PHOEBE ROUGE: Is there a particular setting that if I bought a smart device I should make sure it has, or that I would immediately change when I bought it home?

PARDIS EMAMI-NAEINI: Yeah. So one thing that I'm concerned about, for example, is data being shared with third parties, or my data being sold to third parties. And something that I would look for is, can I opt out from data sharing? And so this is the first thing that I would look for. But as I said, privacy is very subjective, so it really depends on your own preferences.

PHOEBE ROUGE: That makes sense. I guess, Danny, I would ask you the same question. You're looking at all of this data coming out of the smart devices. Is there something specific that you would look for as a control or something you would want to change when you brought it home?

DANNY YUXING HUANG: First thing I want to do when I buy a new device is run it over IoT Inspector and see what's it doing, basically. And just echoing what Pardis said earlier, maybe different people have different privacy preferences. For me, I don't have a lot of tolerance for weird behaviors, but for others, maybe they would be OK with it. So I think having a tool like IoT Inspector allows users to gain transparency into exactly what's going on with the whole network and make a decision themselves, whether to return the product or continue using the product.

PHOEBE ROUGE: Daniel, I'll just ask you the same question.

DANIEL DUBOIS: So usually, the problem is that, depending on a normal consumer, is not able to configure to the privacy settings in the correct way, because usually they are complicated. Sometimes [INAUDIBLE] speed and installing 81 IoT devices, I had trouble configuring some of them. So even if you have a PhD, it might not be enough to do that properly.

So what I do, and I cannot suggest other people, too, unless they have the technical capabilities, is to try to isolate the devices from the public [INAUDIBLE] as much as possible. There are some open source tools, like Home Assistant, that are difficult to use for most people, but maybe in the future, there will be easier versions of that.

And those tools can actually isolate the IoT device from the internet and they can control what the devices are doing and what they are not. And those tools are open sources so they can be analyzed. The code is open for everyone. And if your IoT devices behind do like that, it's much

safer to use than when they use a black box solutions, that you don't know exactly who they are talking to, what they are doing, what they are saying, and everything is like a question mark.

PHOEBE ROUGE: All right. I'll start with Pardis again on this question. So as people become more aware-- we see lots of headlines. We have this whole event, your research and the others getting out there. There's a lot of marketing talk, as far as how much IoT is going to proliferate, and we definitely see a lot of devices being sold.

Do you think, either in the course of your research, as you were asking questions or when you explain your research to others, do you see any changes in people's feelings about IoT, as far as these devices, these clearly require a lot of care and feeding. Do you see people changing their minds and thinking differently about how IoT devices should be used in their home?

PARDIS EMAMI-NAEINI: Great question. Yeah, so in the interviews that we've conducted over the years, we've found that participants are concerned about the privacy and security of smart devices. And they know, for example, smart speakers are very famous. So they know that, for example, they are doing some weird stuff because they've seen that on news, for example.

And so they're very concerned. But at the same time, when you ask them whether they'd purchase the device or not, they would still purchase it. And this is not really about whether they're concerned or not. I think it's mostly about whether there are alternatives in the market, and if consumers know that these alternatives are better, in terms of privacy and security.

So I think there are basically two issues, that you don't really know which devices are better and you don't even know how to define better privacy and security, because at the time of purchase, you have no information about the privacy and security of these devices. So I think if you can solve these two issues, in the market if you can have better products, and if you can convey this to consumers that these are really better products, then I think consumers would be better able to apply their concerns. Now they're concerned, but they don't do anything about their concerns.

PHOEBE ROUGE: Thank you. So one question I wanted to make sure-- to circle back-- we got from the audience, Danny, you had mentioned that your app is not approved for the Mac app store. And I'm wondering, could you just quickly explain why that might be?

DANNY YUXING HUANG: ARP spoofing. It is an attack, basically, but we are turning this attack for good. That's a short answer.

PHOEBE ROUGE: [CHUCKLES] That makes sense. So yeah, we're right up at that time, but I guess I just wanted to give each of you a chance. If there's one thing that you would want consumers to come away with from this presentation and from PrivacyCon, what's one concept that you'd like them to come away with? And I guess we can start with Daniel.

DANIEL DUBOIS: Yes. So one thing that is important to know is that IoT is not going away. It's becoming more common in our lives, so we cannot think that we'll reduce the exposure by just not buying this stuff. You can already see that. Try to buy a TV that is not smart. You will

not be able to find one. And this might become common with many other objects. Of course you will have to connect them to the internet.

In my house, I have a device, a cooking device, that doesn't have any interface on it. It needs at least Bluetooth to work, because it requires a phone. And even if it needs Bluetooth, then the companion app of the device connects to the internet in some way.

So we have to learn how to use these devices properly, and we need to keep doing research on the privacy concerns on them, because regulators will notice when these things are happening. And as it happened already from the apps, the privacy regulations will be updated and the devices will be safer to use, hopefully, and there will be more transparency.

PHOEBE ROUGE: Great. Pardis?

PARDIS EMAMI-NAEINI: So this is not directly related to my presentation, but it's related to the interviews that we've conducted about this study. So I want consumers to know that the smart devices are not pieces of furniture, that you would just have them in your home and that's it and then you don't need to think about them.

Because I've seen a lot of these anecdotes, that people think that-- they're getting used to these smart devices and they don't really care about them. They don't really do anything to change their setting or even think about them. But that is not the case.

These devices are powerful and they will get more powerful in the future. So they have these sensing capabilities. And you should treat them as things or people who can listen to you, or even can see you. And if you treat them like that, you would change your behavior in front of them.

PHOEBE ROUGE: Makes sense. Yeah, and Danny.

DANNY YUXING HUANG: Echoing Pardis's point, these devices are getting more powerful and they're getting more prolific, so what do we do? There's no signs of them improving, in terms of security and privacy, so what we do?

Two suggestions as you walk away from these presentations. One, set up a separate network, just for smart devices. Many home routers allow you to set up a guest network. Just connect your smart devices to a guest network. So increasingly, you're working from home, you probably don't want your regular computers to be talking to and from the smart devices, if they're ever hacked. So one, set up a separate network.

Two, devices like smart TVs, they have capability of tracking. They follow you around. So say, for instance, you don't want to start looking at some shoes on your website and start seeing these shoes in a smart TV, so what do you do? Use a separate account, a separate email address for your smart TV account. For me, I use a completely new Gmail account, just for my smart TV, so that I don't have advertisements follow me around.

PHOEBE ROUGE: Those are good practical suggestions. All right. Well, we went a little over into our lunchtime, but thank you very much for your presentations and the discussion. This was really interesting. And we'll be back after lunch with presentations about specific devices, like cameras and such. So we'll see you back here then.