FTC - PrivacyCon 2022 - Virtual Workshop- November 1, 2022

Jamie Hine:

Good morning. Welcome to our seventh annual PrivacyCon. My name is Jamie Hine. I'm an attorney in the Division of Privacy and Identity Protection, and along with my co-organizer, Lerone Banks, technologist in the Division, we're so happy to have you here this morning. Before we get started, a few details. First, our agenda and bios are located on the event page at ftc.gov, we have our full agenda with links to several of the papers, as well as a list of biographies that you can read. Following the event, we'll have archived all of the presentations today, so in about a week you'll be able to go back in case you missed anything. We'll also update the papers and provide any additional information from today's presentations.

We're live tweeting today's event, so please follow along at @FTC and use the hashtag PrivacyCon 2020... Excuse me, PrivacyCon22. I want to say thank you for all the people that are presenting today. We have 19 people participating. We have 16 presentations and three people with a dedicated conversation panel. We're grateful for all of their time and all the preparations that they made to be here today. Now it's my privilege to turn things over to Federal Trade Commission Chair Lina Kahn.

Chair Khan:

Thanks so much, Jamie. Good morning, everybody. Thanks so much for joining us. PrivacyCon is one of the marquee events that the FTC hosts every year, so we're really, really thrilled to be able to get started today. So, it's no secret that government has not always kept up with the pace of technological change. This has probably always been true to some extent. 18th century constables were probably baffled by the advent of the steam engine. But the digital revolution has been especially challenging. With today's technological revolution, the most important shifts take place out of sight. The surveillance economy became widely entrenched long before it was widely understood, and the law was slow to recognize the harms that flow from it. That's why it's so important that we hear from the academic community.

The commission started PrivacyCon in 2016 to help us get better at spotting challenges before they become crises. Academics have often been the first to see what the FTC and other government bodies may have missed. Calling them Cassandras may be hyperbolic, but only by a little. These presentations that we see at PrivacyCon are often prescient about the privacy issues lurking just around the corner.

If you look at today's agenda, you'll see what I mean. We'll be hearing from innovative researchers across a wide range of privacy-related areas. We have leading experts on topics like ad tech and algorithmic recommendations, and these are areas that are central to data privacy but can be very difficult to understand. I'm also very excited to see that we have panels on emerging fields like virtual

reality and augmented reality. VR and AR are both essentially still in beta without a clear business model, but that hasn't stopped some of the world's biggest technology firms from investing billions. Enforcers and regulators shouldn't wait until a new sector matures before thinking about the issues it could raise, and the FTC is already taking steps to ensure that we're fully abreast of the issues emerging in these emerging sectors before problematic business practices have time to solidify. Listening to the academic community is a key part of how we do that.

PrivacyCon isn't just about looking ahead, it's also about now. At the FTC, we're working on privacy matters constantly as we've been prioritizing the use of creative ideas from academia in our bread-and-butter work. One way we do that is by crafting stronger remedies to reflect what's really happening on the ground. We recently settled a case against alcohol delivery app Drizly, which had suffered two major breaches of consumer data. Our complaint alleged that the company had failed to take reasonable steps to prevent that from happening. The settlement imposes specific requirements that reflect current best practices. For example, Drizly must require employees to use multifactor authentication when accessing consumer data. It must destroy any user data that it no longer needs and document that to the FTC. These might not sound like very cutting edge to true privacy nerds, but they're a big step forward for government enforcers. I'm so grateful to our staff, including DPIP, our Division of Privacy and Identity Protection, for really charting the path forward here.

Speaking of cutting edge, I'm so excited to introduce everybody to our new chief technology officer, Stephanie Nguyen. Stephanie is a brilliant technologist with wide experience across government, the private sector and academia. We're so thrilled to have her. As CTO, Stephanie will be leading the charge to integrate technologists into and across the commission's various lines of work, and she's already been doing this for the last nine months. That interdisciplinary effort will help the FTC be more flexible and nimble in defense of consumer privacy. And you'll be hearing from Stephanie up next. I'd also like to just thank the work of countless FTC staff whose efforts have been integral to PrivacyCon, as well as the folks who really fight to protect Americans' privacy day in, day out.

Today, thank you so much to everybody who contributed from the Bureau of Consumer Protection, the Bureau of Economics, and the Office of Policy Planning. In particular, I'm so grateful to Jamie Hine, Lerone Banks, Molly Smith, Caelan Conant, Jacquelyn Brayboy, for their deep involvement in putting today's event together. Finally, thank you all for tuning in and to all of our researchers and academics and participants who've taken the time to join us today and participate in today's event. We're so grateful that you're choosing to spend your time with us today, and I'm really looking forward to the discussions. With that, I will turn it over to Stephanie.

Stephanie Nguyen:

Thanks so much, Chair Khan, and to the staff whose work led to this conference, especially Jamie and Lerone, who spearheaded this event, Caelan and Molly Smith, and the countless technologist and attorneys for putting in many hours to review papers and plan sessions. And a big welcome to our esteemed experts who will be presenting today. My name is Stephanie Nguyen. I'm the Federal Trade Commission's chief technologist, and I'm honored to be here and to lead and work with world-class technologists and agency staff to make sure the largest corporations follow the law and treat people like human beings. As a human-computer interaction designer and user experience researcher, I've spent most of my career building and designing tech on the ground, working directly with thousands of people in their communities to understand how these technologies impact people's lives. I've seen firsthand how technology can enable or sometimes exacerbate real-life harms. This means patients seeking access to healthcare with the threat of their data being shared to law enforcement or small-town goat farmers who lack the right to choose who can repair their tractors.

In this work, the term consumer privacy is often an inadequate term to describe people's lived experiences. Even looking at our agenda today, PrivacyCon expands on and uses other frames to bring out a broader set of concerns. For example, we're talking about not just indecipherable consent popup windows, but consumer surveillance enabling information and power asymmetries; not just data security, but automated decision-making systems which can bring discriminatory outcomes. Over the past 108 years, the commission has navigated immense change in transformation by dynamically changing our strategic approach, our policy tools, and our operational objectives. I am deeply grateful for the expertise from bureau technologists who have laid much of the groundwork here.

We have learned that tech cannot be viewed in a silo. It cuts across industries, which is why our technologists constantly work with our colleagues across the Bureaus of Competition and Consumer Protection. And this is why our team of AI and security experts, software engineers, designers and data scientists are here to provide diverse skill sets in analytical tools. I want to highlight a few areas of focus for our technologists. First, we strengthen and support the FTC's enforcement efforts with the bureau's staff. On remedies, we are surgically improving our orders to push companies not just to do the minimum to remediate areas like unreasonable data security, but to model best practices we want to see from industry. We want to see bad actors face real consequences.

And to do so, we are holding corporate leadership accountable as we did in our Drizly and our SpyFone cases, and requiring companies to delete the models and algorithms it developed by using the data, photos and videos uploaded by its users, as we did in our Everalbum case. We want to address systemic risks and not play whack-a-mole like requiring companies to delete personal information they illegally collected and destroying any algorithms derived from the data, like in our case against Kurbo, formerly known as Weight Watchers.

Second, we serve as an expert resource to advise and engage with staff and leadership. On AI and complex technical systems, our team is helping demystify hype terms and make sure they don't become obstacles to investigations, like identifying when hiring software that advertises emotion recognition is pseudoscience. We want to help our attorneys be able to interrogate systems and get to the root cause of harm.

Third, we promote best practices of tech policies through outreach and engagement. This may be done through research in horizon scanning, where we aim to establish durable agency muscle memory to stay on development so that the FTC can nimbly identify and respond to current and next-generation tech threats. For example, with augmented and virtual reality, the market has grown significantly in the past half decade to include education, healthcare, and fitness. And as technologists, we ask what, if any, are the novel features that may raise new challenges, like more types of geospatial and biometrics data that can be collected and more inferences made, or more types of content moderation and immersive or invasive advertising experiences.

So, looking ahead, the FTC has and will continue to be measured by the results we can deliver to fulfill our mission and obligation to protect the public from unlawful business practices and from unfair methods of competition. We look forward to a future for technologists to be an institutional resource in the agency and continue this work. Thank you to the team CTO, and to my FTC colleagues. And with that, I'll turn it over to my colleague, Amba Kak, for the first panel on consumer surveillance.

Amba Kak:

Welcome everyone, to the opening panel on consumer surveillance. We use the term consumer surveillance today in its most expanded sense to include workers, of course, and also to leverage surveillance as a frame that creates space for concerns beyond traditional privacy and data security. Our presentations this morning covered very diverse contexts in markets from ad tech to data broker

markets to worker surveillance, but in the Q and A, we're going to try to knit together the common themes that run across these, including the power and information asymmetries that shape these particular contexts, as well as discuss how the research presented today reveals the limitations of dominant approaches, whether that's consent or transparency, or indeed, privacy. And with that opener, I hope that piques your interests. We're going to dive right in. Our first presentation is Piotr.

Piotr Sapiezynski:

Hi, my name is Piotr Sapiezynski, and I am a research scientist at Northeastern University. Most of my work is in auditing algorithms of online platforms, but today I will talk about biases in coverage and accuracy of the data broker data and how these issues can further drive societal disparities. Next slide please.

The FTC has long been interested in the data brokers industry. Already in their 2014 report, the commission noted that the data broker industry is complex and that the actors collect vast amounts of data from a variety of sources without the consent of individuals, and they leave people with little agency over their information. There were two chief worries presented in the report. The first are the inherent privacy risks, of course, of vacuuming and storing all of this data indefinitely. But equally concerning are the situations where despite the wealth of data that the data brokers collect, they still fail to provide accurate information in critical contexts. And in particular, the report presents a scenario in which a consumer is denied the ability to conclude the transaction, say a housing application, based on an error in the risk mitigation product that the data broker provides. That consumer will likely not be aware why they're denied, and they won't be provided with recourse. Next slide please.

And sure enough, this concern soon proved well-grounded. Both The Markup and ProPublica covered stories of individuals that were denied housing, or they were charged unfair insurance premiums because of errors in the data and services purchased from data brokers. Next slide please.

And one particular press story that inspired us to do this research reported that in the beginning of the COVID vaccine distribution in the US, many eligible people were denied their doses at clinics which used data broker services for identity verification. And this was especially troubling because the people of color and those of lower economic status were both at high risk of death from COVID, and at the same time were more likely to be denied the vaccine because of the data broker errors. And so given how important and yet understudied the data brokers are, we decided to shed some light on these issues with coverage and accuracy of their data. But we specifically focused on differences along the lines of race, ethnicity, wealth, and age. Next slide please.

But studying data brokers is notoriously difficult. So, on one hand they do allow individuals to inspect their own data, and this actually led to a number of anecdotal stories in the press where journalists basically make fun of the horribly inaccurate inferences that the data brokers made about them. But on the other hand, as researchers, we want to study these effects at scale, and so inspecting it just our own data is not in us. Now getting access to data or the services of any of the data brokers at scale unfortunately requires talking to a marketing representative, and these people are not exactly willing to allow researchers onto those systems. We had actually previously did. We had actually previously gotten a glimpse into the coverage and accuracy of data broker information when Facebook purchased a big chunk of it, and it allowed advertisers to use it for targeting.

Still, we didn't really know how often Facebook refreshed that data or whether they bought all of it. So, it was hard to really determine how much of the issues that we were observing, both in terms of coverage and accuracy, were at the source at data brokers and how much of this came from how the data was transferred to Facebook. Next slide please.

And so, because of these human gatekeepers not letting us into the data broker systems, we did not get access to actually the identity verification or credit scoring products. So, the study that I'm going to talk to you about today is based on a self-service data [inaudible 00:16:42] product from Experian, which for a short period of time did not require talking to a representative. That service actually allowed anybody to purchase raw information about US residents. And so, we did just that, and then we compared it to publicly available ground truth so that we could answer the question of how comprehensive and how accurate that data actually is. Next slide please.

So, in short, we picked 8,000 individuals from the publicly available North Carolina voter rolls, where the voters self-report their home addresses, race, and age. We made a selection such that we would have a balance of races, genders, and ages, and all the intersections of these attributes. Next slide please. We then hid the race and age information from the voter records, and we purchased raw age information from Experian about each of our individuals. Next slide please.

And so, remember that we are interested in both coverage and accuracy, and this experiment allows us to measure both. Because we know the ground truth information from the voter records about each individual's age, we can tell when the Experian data is correct, incorrect, or missing altogether. And because we bought data that is already publicly available, we did not cost any additional privacy exposure to the individuals in our set. Next slide please.

So, let's take a look at coverage first. For 82% of white voters in our sample, Experian reported having an exact match, which also means that for 18% of registered white in our sample experience, Experian did not find a matching profile in their data. Next slide, please. Now contrast this with the results for Hispanic individuals in our sample. The 27% with no match mean that an average Hispanic person is 50% more likely to not have a matching profile in the Experian data than an average white person. Next slide please. We found that the coverage of Asian and Black individuals was higher than Hispanic of any race, but still lower than those of that of white folks in our sample. But that's just coverage. So, this is Experian claiming that they have some information about the person we requested. Let's look at the accuracy of the age information that we bought. If Experian reports an age that is no more than one year different from the ground truth, we are going to count this as accurate. Next slide please.

We see that the disparities are not only in coverage but also in accuracy. It's more likely that Experian currently incorrectly reports an age of a Hispanic individual than a white individual. So, after this, barely over half of Hispanic individuals in our sample have the correct age in Experian data compared to nearly two-thirds of white individuals. Next slide please.

We also wanted to investigate whether people of lower economic status were more likely to suffer from lower-quality data. We did not know each of voter’s net worth, so we used a crude estimate, the poverty prevalence in the zip code where they reside. We computed the fraction of missing and incorrect data as a function of poverty prevalence. And we see that individuals living in higher poverty rates are more likely not to be covered by Experian. And even if they are, it's more likely that Experian would report incorrect age about them. This generally hold for all races that we looked at, but you can see from the slope of the red line that the effect is strongest for Hispanic individuals in our data. Next slide please.

So, remember that we used voter records as our ground truth. That means that the problems we observe affect adult citizens with information publicly available about them. We only purchased one straightforward bit of information, the age of individuals, not the more refined credit information or purchase behavior insights. Because of all this, you could argue that this is an overly easy benchmark and yet the company manages to fail in so many cases. If this data is a base for services like credit or identity verification, it will cause failures that can further disadvantage the populations of color and those more likely to be in poverty.

Here I'll also mention that the Experian did review our research. They claim that the data that they use for sensitive services is completely separate, and we cannot draw any conclusions from our research. In return, we of course asked to be allowed to perform the same study on identity verification data, and then we never heard back from them again. Next slide please.

So, whenever problems of bias are surfaced in machine learning products, there are calls for better and more comprehensive data sets. But I really want to be clear here that we are not arguing that data brokers should be forced to collect even more data about minority individuals. Right now, there are barely any restrictions on data collection, and yet our broker failed to provide the most basic service, and they did so in a discriminatory way. More transparency would certainly be welcome given how difficult it is to study this system, but it is not a solution on own.

We already know, both directly from the stories of affected individuals and at scale from our research, that relying on data brokers leads to discriminatory outcomes. But perhaps we wouldn't need identity verification as a service in the first place if access to real ID was equitable. Perhaps landlords and employers should not be able to base their decisions on credit history checks on the applicants. And so instead of more comprehensive data collection or another transparency initiative that allows data brokers to claim that they're doing their part without actually changing anything, we would like to see systemic changes that allow companies, healthcare providers, landlords, and others to decrease their reliance on data broker services. Thank you very much, and I'm looking forward to the discussion.

Nazanin Andalibi:

Hi everyone, my name is Nazanin Andalibi, and I am here today to share some work around the implications of emotion AI for the future of work. I will highlight two studies forthcoming in an ACM venue, and I will have free PDFs of both of these studies on my website soon. Broadly, emotion AI or automatic emotion recognition technology refers to technologies that claim to algorithmically infer emotions, moods, and other affective phenomena from all sorts of data like the face, audio, texts, heart rate and more. It is increasingly used in a range of domains like the workplace, hiring, education, law enforcement, healthcare marketing, automotive industry, and more. Examining emotion AI's implications is important because people desire control and privacy over their emotions and emotion data. Emotions provide insight into behavior and shape human experience, decision making and wellbeing.

And while emotion AI is critiqued for its validity, bias, and surveillance concerns, it continues to be patented, developed, and used without much public debate, resistance, or regulation. My research group investigates emotion AI's implications in a range of contexts like healthcare and workplace, for example, by examining patents and the perspectives of people who are or would be impacted by emotion AI, some of which I will highlight in this talk.

Oh, I should have said next, but I didn't. So we should be on the slide that says patents right now. Sorry about that. So first we qualitatively analyzed patent applications that developed emotion AI to be used in the workplace, and there were a total of 86 of them as of 2021 when we collected these data. And we wanted to essentially reveal their ethical and societal implications. While patents are legal artifacts evaluated in the US based on criteria including novelty, usefulness, and non-obviousness, according to science and technology scholars like Shobita Parthasarathy, patents are also political artifacts. I see patents as artifacts that allow us to reflect on technology's ethical implications and our sociotechnical futures as well as ways that we can intervene in potentially harmful futures. With this study, we asked what input data do these technologies described in these patents use, what outputs do they claim to infer, and what actions do they take or attempt to prompt, and how are these useful?

So, in the interest of time, I'll share highlights about both kind of what we found about input data and usefulness. On input data, the patents use a wide range of input data, most commonly text, speech, facial images, non-facial biometrics, physical activity, computing behaviors of targets, contextual data and more. For example, this patent describes on emotion AI that uses data like speech, computing behavior, posture... I'm sorry, I forget to say next. So we should be on the slide that says input right now. So, for example, input with a large list of input data types.

For example, this patent describes an emotion AI that uses data like speech, computing behavior, posture, and biometrics to detect the comprehensive state of workers, interviewees, and customers, and later indicates that the technology can use all these input data that I listed on this slide. I won't read through all of them, but I hope that you are overwhelmed by looking at the list.

I mentioned this example because while facial emotion recognition was a common type of emotion recognition... I think there's something up with the slides. So, we should be on the input slide that lists a series of input data types. I mentioned this example because while facial emotion recognition was a common type of emotion recognition present in our analysis, and while I echo efforts to resist facial recognition and facial emotion recognition's harms, we identified such broad input data types and language across these patents. This common approach, given that these are patents, might be a way for an applicant to claim a broad range of appropriate data collection in case they want other competitors to not use it or to conceal what their real purpose is.

Regardless, in practice, this approach also leaves space for extremely broad and invasive data collection and is aligned with broader computing approaches aiming to collect more and any data. This pattern is problematic as it may also support employers, technologists, or other actors to move away from the collection of face data and simply shift to collecting other data types, such as those that we see here, without truly addressing the fundamental harmful implications of using these data to infer emotion and affective phenomena. Next slide please.

So, we also look to see what benefits these patents claim that their technology has. As noted earlier, patents are evaluated in the US based on several criteria, one of which is utility or usefulness. So, most benefits were framed to be for employers, but some also claimed benefits for employees. On the left side of this slide, we have claimed benefits to companies, employers. Overall, there's this theme of a desire to make the concealable visible without choice and conflating privacy management at work, which people have been doing forever, with deception as something that needs to be addressed here with this technological intervention of emotion AI. On the right, we have claimed benefits to workers such as improving mental health or providing support to workers who might be struggling. While these are examples of claimed benefits to workers, it is unclear if workers would agree that these would be beneficial to them.

So, in a preliminary survey with about 400 people, we explored people's attitudes towards emotion AI at work. Next slide please. So, we found that around 32% of participants perceive no benefit to them, not at all. So, sentiments like, "I think this would give employers a tool that could really help them but not the employee," they're pretty common. Next slide please.

Second, while some noted potential benefits, they still anticipated several harms. They expressed concerns regarding the potential for emotion AI use to harm their wellbeing, privacy, work environment and employment status, and to create and amplify bias and stigma against them, especially the most marginalized along dimensions of race, gender, mental health status and disability.

These findings, which I won't go through in detail here, reveal how emotion AI may indeed magnify rather than alleviate some of the existing challenges, like worker wellbeing and worker stress that work is already face in the workplace. Next slide please. For instance, this participant who disclosed having multiple health conditions shared concerns regarding negative wellbeing implications as a result of

emotion AI-induced privacy intrusions. They say, "The awareness that I'm being analyzed would ironically have a negative effect on my mental health." Next slide please.

Another participant's remarks demonstrate concerns around how employers using emotion AI could potentially discriminate against marginalized workers. This participant says, "These systems have the potential for both racial and gender biases, particularly against people of color, women and trans individuals. Who is deciding what expressions look violent? A system can read faces, sure, but not minds. I just cannot see how this could actually be anything but destructive to minorities in the workplace." Next slide please.

So, some takeaways. First, emotion data is sensitive data, and we should be wary of moving from the collection of face data only simply to shift to other harmful data collection, such as those that we uncovered in this patent analysis, without truly addressing the fundamental harmful implications of technologies that use these data. Second, workers' perceptions of emotion AI in the workplace, as we examine, challenge the dominant discourse surrounding emotion AI and its stated purposes, for instance, purposes to improve worker wellbeing or work environment that we also saw was present in our patent analysis, showing how emotion AI use could ultimately harm the very conditions that proponents of emotion AI claim it will improve.

Third, some scholarship and practice aim to improve emotion AI for accuracy and bias. This work suggests that even if emotion AI's technical inaccuracy and bias concerns are somehow addressed in the future, there are still emotion AI-inflicted harms that increased accuracy may indeed exacerbate or that devised systems would not actually mitigate, such as privacy and wellbeing related harms. This raises the larger question of whether emotion AI can ethically be implemented in the workplace at all.

So, with that, I want to emphasize the choice that designers have to not build emotion AI and for organizations to not deploy it. And finally, as workplace surveillance increases its reach beyond behavior and into our internal states, it is important that regulators provide protection for impacted groups like workers, and consider adding additional scrutiny to patents as well as resulting services, and more broadly, our innovation evaluation process that foregrounds their societal impacts beyond our

Nazanin Andalibi:

... The current criteria of novelty, utility, and non-obviousness. With that, I'm more than happy to connect with folks one on one about any of these studies and some of the other work in this space and I look forward to the conversation.

Patrick Parham:

Hi, I'm Patrick Parham. I'm a PhD student at the University of Maryland. My research primarily focuses on privacy and advertising technology, and today I'm going to present a paper that I've authored with my advisor, Ido Sivan-Sevilla. And this paper addresses the current state of cookies in the advertising industry and the network effects that they create through supply side ad networks, and then the current state of proposals in what is called the cookie-less future solutions for tracking and identification of users, primarily based on IDs created through personally identifiable information. With that, I'll start. Next slide please.

First, in our study we build a novel typology to first specify what we... Okay, so the next slide please. Yes. The research design and methodology. Looking at cookie-based tracking specifically, we identified four SSP networks to perform our analysis and scraping of the web and generated a list of 50 popular websites. And in that analysis, we performed 10 individual stateful crawls over these 50 websites. And we looked at the persistent identification of cookie IDs across these sites. And to specify all sites worked with all four SSP networks to present some level of consistency. And then looking at cookie-less

proposals, we analyzed technical documentation, company statements, and then trade publications of coverage of three solutions, specifically the Trade Desk primary DST, their solution, the Unified ID 2.0, the LiveRamp encoding solution, and then the SWAN proposal. With that, I will ask for the next slide.

And then in terms of how we define tracking with cookie-based solutions and cookie-less solutions, first the user identification instrument. We define that as just the pure mechanism of how an individual is identified. And then cross site usability, are they able to be seen across publisher websites? Longitudinal tracking, are they able to be seen across a period of time? Circumvention of targeting restrictions, are they able to evade certain sensitive categories that platforms such as ESPs restrict? And then user data sources, are they able to also incorporate data from other third parties that might enable this circumvention? With that, the next slide.

When we think about how we're profiled on the web, our identity is somewhat fragmented despite what we might think about a consolidated view that advertisers might have. The current view in our mind of network effects is that yes, you were profiled as you visit multiple sites, but those profiles are siloed and partitioned based on the ad network or supply site network that drops a cookie to identify you. In our crawls, OpenX might have a view of you, AppNexus might have a view of you, and Rubicon and PubMatic might all have a separate view of you. But if we go to the next slide, in the cookie-less future, these profiles would be combined as your PII would generate an ID and you'd be identified across supply-side networks and your identity is further consolidated into one profile.

Next slide please. To give an understanding of the level of persistent identification in the current cookie-based landscape, when we crawled the four SSP networks and the 50 popular websites, on average users were identified on publisher websites 86% of the time across before supply side networks. You might think this is high or low, but the drop off from 100% to 86% was primarily caused by first a cookie not being loaded on the page potentially, and the lack of clarity between what is included in a publisher's ads TXT corresponding who they're able to work with and then the reality of that as that is not actually up to date in all instances. And with that I will say let's go to the next slide. And to compare finally SSP cookie-based tracking within their networks to universal IDs.

Looking at the user identification instrument, cookies as we know are just the passive placing of third-party ID cookies when you visit a page versus universal IDs or the cookie-less solutions where a cookie is in the background for first party cookie. But the identification is primarily based on consent obtained on the publisher's site and the sharing of personally identifiable information, in most cases the email address as you're logging in to the publisher's site. And as we all know, this is generally not the best mechanism to prevent privacy issues as the individual or the user will most likely opt to share their information to just engage with the content. Then in terms of cross site user visibility, as we demonstrated there's 77 to 90% coverage of persistent identification of individuals across these networks and individual sites. And it's our view that this will be further consolidated as this is consolidated within an individual profile and is shared across these four SSP networks and it's expanded to be a consolidated profile of the individual and is no longer limited to these partition networks.

And then longitudinal tracking, yes there is tracking of individuals across time with cookies, but those can be refreshed and deprecated. But also, with universal IDs there's more reliance on deterministic data, so it's further tied to a particular piece of information about you over a longer period of time. And then the circumvention of targeting restrictions. Yes, we know that third party cookie data can be paired with data broker data, but in the universal ID based solutions, there's a further importance placed on publisher first party data as that fuels the network and the advertising industry going forward. But what that also means is that there's a fueling on the other end of the advertiser side where first party data can be enriched and further expanded on. And that encourages the pairing of first party data with third party data purchase from data brokers. And while advertisers might be limited on the platform side by

certain restrictions set by the platforms, they're able to manipulate data beforehand before engaging with the platforms.

Looking at user data sources, it's unclear still with cookie-less solutions how they're able to circumvent that, but previous work has demonstrated they're able to do so rather easily. And then next slide. In conclusion, based on what we have presented, we first argue that persistent identification has actually increased in a cookie-less future based on the fact that identification of users is no longer limited to partition networks being the SSPs and then the source of identification is no longer determined by third party cookies, but it's determined by PII. And so that longitudinal component makes the persistent identification take place over a longer period of time and is actually more difficult for users to opt out based on the consent-based mechanism implemented here. And then third, the importance of PII as an alternative to identify users as encourage advertisers to purchase first and third-party data to segment audiences before interacting with primary platforms and create rich profiles of users that incorporate traits that violate the policy of platforms and prohibit the targeting of consumers on sensitive categories.

If we go to the next slide, so for future work and interest going forward primarily relate to this third component where we think there's a large amount of research that needs to be done and can be invested in understanding how the shift is taking place from going from the platform level to the individual advertiser level based on this disconnect between platforms and then how advertisers are being incentivized to target individuals going forward based on the importance placed on first party data and enriching that with third party data purchaser data brokers. While advertising platforms create these policies, they're easily circumvented and they're really loosely enforceable and there's no mechanism currently in place to actually verify on the platform and what is going into the pools of IDs that are being pushed into platforms.

And at the end of the day, we believe that with this current emphasis in the landscape, it's just pushing the issue from the platform level, just pushing the liability down the ad supply chain. And with that, I'm looking forward to any conversation we have and any questions anybody might have and if anybody wants to email us for any comments, we would really appreciate that.

Dan Calacci:

Okay. Hi everybody, I'm Dan Calacci. I'm a PhD candidate at the Media Lab at MIT under Sandy Pentland. And today I want to talk to you about worker data and why I think worker data regulation really shouldn't be entirely about privacy. I'm going to do that by talking about some work that we've done over the past couple of years. Next slide. Next slide please. Great. This is Willie Salise. In 2019, Willie signed up as a gig worker for Shipt, Target's delivery company, after losing his job as a construction worker. At first, working for Shipt was great for Willie. For each order, Willie got a $5 base pay plus 7.5% of any customer's order on goods from Target. But then in 2020 something changed. On Shipt Facebook groups and forums, Willie noticed posts complaining about pay changing. Shipt shoppers suddenly weren't getting paid that clear fee anymore. In fact, shoppers couldn't really tell why they were getting paid what they were getting paid at all. This is because Shipt had started to move to a black box algorithm to pay their workers. Next slide.

After talking to hundreds of workers, Willie decided to answer the biggest questions that folks had about the new algorithm with data that Shipt wouldn't provide. Willie started collecting screenshots from workers to document any changes to pay that had happened over time and workers would text Willie these screenshots and really would then put them in a spreadsheet. Next slide. And he started getting hundreds of screenshots from workers, way too many for him to really manage on his own. He got so many that to document all of them, he really had to stop working and focus on this data

collection full time. And this was to answer these questions about this new algorithm. If the new algorithm was different from the first one, how it impacted people's pay, who was being affected? Is it impacting certain places or certain kinds of workers, certain kinds of consumers over others?

And to answer these basic questions, next slide, we worked with Willie to develop a texting bot to automatically do the work that Willie started, engage the community, and create a much larger data set that we could use to definitively answer some of these questions. This tool that we made also crucially gave workers information that the company refused to. This is a common request we got from workers while co-designing the tool that they wanted to know individually whether they were being paid by the new algorithm, how much their earnings had changed, and other useful information like the percentage of their pay that came from tips. This is all information, mind you, that the company has but doesn't disclose to workers through the interfaces that they develop. And so, in order to elucidate this, workers have to collect it themselves or work with us to collaborate on tools to actually find it.

Next slide. What we found with this data set after auditing Shipt's pay change is that it cut the pay from over 40% of the workers who participated in the project. Organizers used our report that we developed after analyzing this data to organize campaigns and plan work stoppages to put pressure on Shipt. And these pay changes are not just one-off jobs. There's entire categories of workers who suddenly got a large pay shift without notice, without announcement because of this algorithm changed. And this is just one change and one relatively small platform. Major platforms make these kinds of changes constantly. Gig workers frequently speculate about new versions of algorithms that impact their pay and their working conditions, but these platforms are opaque, and workers lack access to that data and the resources even needed to collect it or analyze it, so they're left speculating online.

Next slide. This project has since been covered by NPR's Radiolab, Mozilla's Internet Health Report, other venues. And we presented it as a keynote at a conference in algorithm fairness in Seoul, South Korea this year. And this one small example where workers actually got the data capture journalists and academic attention is its indicative of something. I think this is illustrated for two main reasons. The first is this is not just limited to gig work. Next slide. This, once it comes up, is a mouse jiggler. It automatically moves your mouse while you're away from your computer. Now this is not directly related to Willie's work, but office workers who are subject to similar kinds of algorithmic management where their quantified and their productivity is measured. People who started working from home have started buying these devices because they found that under new algorithms that measure their productivity, they'll be docked for being away from their computer for a bio break or a stretch.

Next slide. Office workers in more traditional employment relationships are increasingly subject to these same kinds of opaque monitoring and scoring systems similar to the ones that Willie and other gig workers are facing. This means that getting data protection at work right is crucial for the future of many more people than just gig workers, even as gig work is ballooning. Next slide. I want to go through some of the lessons that I think we've learned in thinking about the future of data regulation at work from the project we did with Willie and thinking about this more broadly. Willie's project and the mouse jiggler both point to an important underlooked aspect of data protection at work specifically. Although platform workers are really heavily surveilled, Willie wasn't collecting his fellow workers' data to protect privacy. They wanted to access their data, they wanted access to the data that was collected from them by the platform to better understand their conditions at work in order to organize, build worker power, have alternate narratives from the employer's story and hold their employer to account.

But almost all data protection regulation either has a legal basis in privacy rights or really myopically focuses on the individual data subject. And this really limits data protection's role in building worker power and actually working for workers and filling workers' needs. The only exception here is probably the recent EU platform worker directive, but even that piece of regulation doesn't include access to data

like the information that Shipt workers got here through this project. Next slide. I'm just going to go through a few broad areas that we think are important to consider in the future of data regulation at work. Another one is that the central goal of any regulation that addresses worker data, especially for platform workers, should really be to reduce information asymmetry between the platforms and the workers. Less about privacy, more about access to data in order to reduce this power differential. Because this information asymmetry is really what hurts workers the most. Not knowing about a platform pay change impacts their freedom to make an informed decision about what platforms to work on, about their own financial decisions outside of work.

And it allows companies to take advantage of workers constantly to adjust that pay to the lowest amount that a worker might accept to make pay changes like the one that Shipt did every day or every week. Next slide. Our project with Willie also shows that even simple audits like the one we did here, which is a quite simple audit still requires significant resources without additional infrastructure. This is a big obstacle for workers who are trying to understand the algorithms and systems that define their work. Workers need third party actors like MIT researchers to collect and analyze their information, but this process shouldn't really be the burden of already precarious workers or if it is they should have the materials and resources they need to do similar kinds of auditing practices. Next slide. There's a question which is how this actually translates to meaningful regulatory goals. Next slide. I'm going to go through three potential avenues specifically related to the FTC's role. As other scholars have argued, mandated disclosure guidelines could be required by employers through things like the disclosure rules within the franchising enforced by the FTC there.

Now there are legal arguments for considering platforms in particular franchises. The FTC could leverage that and consider data sharing and access a limited disclosure of information that's relevant to consumers. But beyond providing data access to workers, companies could be required to perform audits similar to the one we ran with workers to create more effective disclosures similar to algorithmic impact assessments. And these should be regulated as to be understandable by the average worker or consumer. The length and complexity of disclosures within franchising is notorious as being way too complex and difficult to understand in a way that actually limits people's access to their rights under franchising regulation. Next slide. But one issue with just disclosure as an instrument is that it wouldn't really necessarily involve the level of detail that we collected or analyzed in our report in our work with Willie if they're just made public.

Trade secret IP law could limit what companies can disclose publicly. They want to protect their own property. Disclosure as of yet doesn't include the actual data produced by workers in the employment relationship. But something like a private disclosure rule between parties, individual workers, or worker representatives, could work to limit how that disposed data is used while still providing meaningful access to workers. Next slide. And because data use and data access can't really be separated from issues of working conditions, that's one of the major lessons here is that the algorithms and data that are part of your work and that your employer subjects you to are really issues of working conditions, not just of data protection and privacy. Any omnibus regulation can really risk falling short because it can't cover all areas of harm and risk involved in surveillance and data use. And so, these tools, these disclosure ideas all provide workers with greater agency, but there are limited recourses that workers can use to take the information they learn from that disclosure to exert power in the workplace.

But if worker data use and collection could be regulated through a firm or a sector-wide civic data trust structure, which plenty of researchers are at least looking at these days, workers could exert some control over how their data is used by their employer. They could set limits on what data could be used to create productivity scores or assign work or be used for emotional AI inference. This could facilitate meaningful private bargaining rights between worker groups that are represented by a trust and an employer that could regulate how data gets used and how algorithms get applied to workers in these

smaller scales beyond omnibus regulation. Next slide. And so finally, all these rules apply not just to workers but to consumers. I'm really happy to see a move looking at how commercial surveillance really does impact people at work, not just consumers who are using web browsers. And I'm looking forward to future regulation that helps workers gain more agency over their data at work. Thank you.

Amba Kak:

Great. I think with that we can close the round of presentations. Thank you all for your excellent presentations and we're doing pretty well on time, so that leaves a good amount of time for the discussion. Just to start us off, I think I'll start with specific questions that follow from each of our presenter's research and then we'll follow up with a round of crosscutting themes and bigger, more existential questions for the field that I'm personally very excited to ask our panelists. Just to switch up the order maybe because Dan's presentation was last. Dan, I'll start with you. You mentioned specifically the rise of worker productivity scores in offices. These are sometimes referred to as the white-collar context in addition to platform or gig workers or say warehouse workers for tech companies. Now there's been a heated debate for many years around the risk of privacy becoming a luxury that only the well-resourced and privileged in society might enjoy. I was thinking in the context of worker surveillance, where do you see opportunities for solidarity and conversely maybe challenges with unifying these diverse contexts?

Dan Calacci:

Yeah, it's a great question. I think I'll answer it in two parts. First, I think that while certain kinds of privacy are luxuries for privileged workers either in part because of their ability to have more choice over where they work, this is changing. White collar work is becoming increasingly surveilled. And algorithm data are being used within white collar work to do all sorts of management techniques that gig workers and lower wage workers have been subject to for a decade. This is changing and the relationship between who is surveilled and bottom level of surveillance that you might be subject to at work is really increasing. That floor is coming up.

And so that actually offers opportunity for solidarity between classes of workers because there are these same fundamental needs in how people might seek to use data at work because all of these systems fundamentally impact worker agency, their ability to represent themselves in the workplace and to have a say in their working conditions. And so, in all of these cases, I think the hope is that any regulatory tools or even technical tools that might be used in one case could be translated to another because those issues are so similar in terms of agency over your data, how it's used, what algorithms you want to be subject to, how the algorithms work. And so, I think the big hope is to create solidarity through a shared understanding of the impact of surveillance and data use.

Amba Kak:

Thanks, Dan. And while we're in worker surveillance, Nazanin, I'll move to you. At this point, I think there's a volume of research on consumer attitudes to privacy, which has been leveraged towards a spectrum of policy arguments. On the one hand we hear consumers don't care about privacy and therefore we don't need laws at all. Or we might find that they're arguing that consumers do find certain inferences creepy, and that becomes a reason for cracking down. Less so of course on worker or even employer perspectives of the kind that you highlighted in your presentation. I wanted to ask where do

you see both potential and pitfalls of using empirical work on worker attitudes and experiences and how those might contribute to the policy debate on worker surveillance?

Nazanin Andalibi:

Yeah, thank you for that question. A few things. One is that workers concerns with emotion AI in particular include but go beyond the creepiness and privacy concerns, resonating a little bit I think with Dan's presentation. I think for instance, one type of harm that isn't captured quite within a privacy framework is emotional labor. What we see in our work is that, so there we know that there are certain emotional labor expectations within the workplace, and these are disparately distributed. For instance, women, disabled people, women of color tend to have different emotional labor expectations within the workplace to comply with or different occupations like service industry workers or call center workers. There's a lot of emotional labor expectations there. What we see is that emotion AI can function to enforce workers' compliance with normative emotional labor expectations in the workplace on the one hand.

And then on the other hand, workers may engage in these emotional labor practices to protect their privacy. It's intertwined as far as the impact of emotion AI goes in relation to privacy, but it's really beyond that. I think the takeaway here is that there are other harms beyond privacy related harms that people would experience as a result of emotion AI or more broadly, I would argue, workplace surveillance that would need to be addressed within any sort of policy and regulatory framework. The other piece I wanted to say is that one thing that I think is pretty critical within the workplace context compared to more broadly consumer "protection" is the deeply rooted power imbalances between workers and employers.

In that sense, what worker chooses to work at what place or don't work at what place and the amount of choice that people would actually have is very limited. To make things a little bit darker, I would argue that much of the harms and practices that using emotion AI in the workplace would enable also work to reinforce the existing power imbalances that workers already live within in. There is that. And finally, I think it's important to note that some workers within the workplace surveillance context, like other surveillance practices, some would find it more harmful than others, some would in practice experience more harm than others, depending in this case, for instance, on people's social positions, on identities. And so, keeping these kinds of disparate impacts, discriminatory impacts of these technologies would be very important in thinking through different policy approaches. I think I'll end there.

Amba Kak:

Thank you. I'm going to move to Patrick. Shifting gears a bit, moving to ad tech. Patrick, the debate on third party cookies and the deprecation of third-party cookies, including with regulatory moves like Privacy Sandbox has really brought to the fore the intersection and also the tensions potentially between privacy and competition. Your research underscores the counterintuitive privacy harms associated with what you call the cookie-less world. I was wondering if you had thought through potential consequences, if any, of this cookie-less world from a competition perspective.

Patrick Parham:

And I wanted to clarify one point in the presentation regarding SSP selection. The four SSPs that we selected for the quantitative portion of our study have all partnered with the Trade Desk Unified ID 2.0 solution. And looking at the competition aspect of this, currently there are hundreds of SSPs that sites partner with to sell their advertising and inventory and different partners buy exclusively from or put together a package to buy from. And that's the main mechanism for third party placement of cookies

and identification of individuals. Going forward, in our mind, the identification of individuals becomes more just technical, it's no longer competitive. These networks, while they're partnering with these ID solutions, it's not clear what their value is in this exchange. And we've even seen in some of the documentation the ability for publishers to directly integrate with some of these solutions.

It's unclear how these supply side networks will have any value going forward. But the competition here between SWAN, Trade Desk Unified ID 2.0, and LiveRamp or any other solutions all arise is how you can cultivate certain partners to adopt your solution, but also getting your own network going of replicating these supply side networks within a larger pool. And certain publishers have already said, we want no part of these ID solutions, and we'll just sell our advertising directly. There is definitely a competitive aspect going forward and some people are choosing not to engage with it.

Amba Kak:

Thanks, Patrick. Finally, I'm going to move to you, Piotr. You identified data brokers as maybe a particularly troublesome actor given that they operate with structural opacity. I guess I wanted to ask, do you think that transparency is a fix for many of the challenges you uncovered? Or maybe put differently, if data broker practices were made fully transparent, what problems would be fixed, and maybe which ones would still remain?

Piotr Sapiezynski:

Yeah, great question. Thank you for this. In my work, I mostly see transparency as a means to an end, not a solution to any particular problems, but a way for us to better understand how these systems work and how these systems fail to work. And so, I of course understand why they would want to limit that transparency such that it's more difficult for us to determine in which ways these systems fail. But I also want to stress here that... As I said, transparency would be a means to an end. And the more important questions would be whether we actually want these systems in place, and if we do decide that we really need them and we really want these systems, then it's a way for us to verify whether these systems actually do what they promise to do. And so again, this is just a way for us to examine the systems more than a solution to any particular problem.

And this is the case with data brokers here. The problem of identity verification, in other places, identity verification is just solved by using national IDs. And then you don't need third parties to do all of this data collection analysis and selling this as a product. If people had easier access to IDs without maybe requiring taking a day off and then going to a remote location to get that ID, and then refreshing this with quite a lot of money every few years, maybe we wouldn't need that product whatsoever. So, I think the question of, then again, even bias or transparency, these are secondary to asking, do we actually need this? Is there an easier solution that doesn't require collecting all of this data to achieve the goal that we're actually trying to achieve here?

Amba Kak:

Yeah, that's really well put. I think now I'll move to kind of the cross-cutting questions. And I'd love for as many of you that have thoughts to come in on these. Maybe sort of related in some ways to the transparency question, you've all gestured in your presentations to access to data challenges and they really kind of underscore the power information and any resource disadvantage for those scrutinizing these systems on the outside, whether that is researchers or that is advocacy or policy organizations. I guess as researchers, what is your take on how these kinds of challenges might be surfaced and the more kind of existential question if and how they might be overcome and... Yeah, I guess anybody who wants to go first can start, and if not, I will point to somebody. Piotr, want to go?

Piotr Sapiezynski:

Right. I think... So, as I mentioned in the beginning, most of my work is actually on online platforms, not necessarily data brokers. So, this is the problem that we have been dealing with in a slightly different context. So, most of my work is on the advertising platform on Facebook. Most of our experiments over the last few years were exactly to bring more transparency to how these systems operate so that then we can bring accountability. One part of it is that even though the systems might seem obvious to people who actually work on with them, I think that there is a big disconnect between experts who find most of our research obvious and people who are really outraged by the findings of ours. And so, I think this is about bridging this gap and translating things that might be obvious to experts to society at large. And I think that this is a way then to influence regulation and help push for better solutions at scale.

Amba Kak:

Dan?

Dan Calacci:

Yeah, so beyond working with good workers, some of my other research also involves examining the privacy implications of big changes that companies make to certain systems. I think one of the really important things to recognize, especially when talking about consumer regulation are protecting consumers using regulation is the extent to which changes on something like Facebook's ad platform or Google proposed a change to their ad tracking ecosystem last year that was eventually canned. Now they're working on something new called Flock that can make fundamental changes to the way that we interact with web infrastructure. At a certain scale, these platforms just become the way that people interact as consumers with every business you can think of or each other. And these changes are often really opaque.

This is true at the micro-scale when you're a platform worker and the algorithm shifts, but it's also true at a really macro-scale when Google changes the way ad targeting works and now, Flock, we found could significant privacy information about its users if analyzed in the right way. But in order to do that analysis, we had to buy a separate data set. It was expensive. We had to do a lot of work. It took a few months of work by two MIT graduate students to find these privacy risks.

And we were only able to do it because we had access to this data set. Google wouldn't release a data set that we could use to analyze their new tool as a third-party. And so, I think at a certain scale, some of these companies should be required to disclose or release tools that allow third parties to audit how these changes might impact the consumer ecosystem. So, for example, if Google proposes a change that's going to impact 60 million users through its Chrome browser, I think that they should be asked to disclose or share even a toy data dataset that people can use, or a toy algorithm people can use to examine its potential impact.

This could be governed through... I've already talked about data trusts once, so I'll just drop it one more time. This could be governed through something like a data trust that provides limited access to researchers who would have to make an agreement with, say the FTC houses it, I have to make a data use agreement with the FTC that I'm using it for this specific auditing purpose, but I can publish my findings. That would make access to the information needed to do these kinds of audits way more accessible to many more researchers than just myself and would help turn transparency into something that can actually turn into action.

Amba Kak:

Nazanin?

Nazanin Andalibi:

Following up on that, I wanted to say two things. One, with the emotion AI and hiring work that we've been doing, I actually tried at some point to... And I had money, it wasn't even a concern, but kind of do I have resources. I had money. I wanted to essentially just get into these systems, see what they do and kind of use them as a researcher and run experiments and such. It was impossible. The amount of information that they asked me to provide them with, all of these services, just pretty much made it impossible for me to do it. I stopped it. I'm pretty sure there are probably some workarounds, but I shouldn't have to think of all these kinds of magical ways to be able to really see what are these systems doing? What is the experience on the applicant side, for instance?

So that's one thing. And then the other piece I wanted to say is that we looked at some... Within the same context, we wanted to examine how do emotion AI, emotional recognition service vendors within the hiring context, kind of legitimize their existence, what do they say the benefits are, what problems are they solving, and that sort of thing. It was extremely difficult to locate and identify these services, and that is partly because the vocabulary that these services use is extremely inconsistent and not accurate with what they actually do. So, for instance, you might see, "Oh, we use algorithms or AI or analytics for machine learning to identify the best candidates for your position or to find the best fit for your job," or to identify soft skills or the types of things that companies tend to care for using some form of magical technology. And so, you can't really find these with keyboard searches or anything that would actually resemble what the technology actually is, which is deceptive.

And so, it took us months to be able to really go through all of the vendor services that claim that they use some form of technology to go into their material and see, okay, what is it that they're actually identifying? What sort of information do they have available to the public about what is it that they do and what forms of technology they're actually using? So, I think this is more of a... I guess, expressing what sorts of problems it causes for researchers rather than providing solutions. But I just wanted to highlight that as one example of it being extremely difficult to actually see what's going on.

Amba Kak:

I appreciate that and maybe the difficulty with coming up with solutions is also part of the diagnosis here. So maybe to close us out, I had one big question, which I hope maybe all of you could come in on. So, we talked about in the beginning that dominant approaches like consent, privacy bias, they do shape the way that we think about these particular concerns, and then also the imagination of what we might do to remedy these problems. One thing that struck me across your presentations was the kind of challenge to some of these frames. Nazanin, you challenge maybe bias as too limited. Dan, you called out privacy as being an unsuitable or narrow frame for worker surveillance. And Patrick, your presentation kind of explained how the cookie list consent-based world was not necessarily a privacy improvement. So, could each of you maybe speak to that and what you all see as compelling alternatives? Nazanin, I'll start with you. So, if not bias, then what?

Nazanin Andalibi:

Yeah, I think that's a great question. So, one thing that I see in my work is that even, as you mentioned, even if issues with bias are resolved, there are still other harms. Even if privacy, even if people aren't concerned about privacy for some reason, there are still other harms. Consent is pretty much out the door when they are in a power imbalance situation such as within the workplace. So that is that with

these three frames to be brief. But I think one of the limitations of these frames is that they attend less to the processes than mechanisms through which people experience harm.

And they also make it difficult to determine both short-term but also downstream harms that can result from these technologies at multiple levels. So, at the individual level, let's say a worker, interpersonal level kind of worker-relationship with the employer and institution. At a societal level, what sort of society are we building?

So, I think that's... To me, I think that this is kind of the limitations that these frames have. So, I think frames that enable us to more holistically explain harmful processes and harmful outcomes are a reasonable way to go. And to me, theories of justice or a justice frame is really powerful because it allows us to... It kind of combats these limitations and that allows us to both understand unjust and harmful outcomes, but also processes and how these harms come about to be able to find ways to intervene at different stages of processes that lead to harms.

And with that, I think it's important to always account for the different factors such as social and identity positions that shape people's experiences with these technologies, including in the workplace, but really also beyond. And theories, for instance, or frames of intersectionality, for instance, allow us to tease out what are the ways that different groups of people experience different sorts of outcomes. So, I think a combination of, at these sorts of frames, seem to be more powerful to me than our existing ways of looking at these problems.

Amba Kak:

Patrick, you want to go next?

Patrick Parham:

Yeah. I think we are pretty blatant in our presentation about consent, how this is just like a formality and the background of tracking is pretty much the same in a cookie-less future. But in terms of our hope of how things could ideally be done, I think those publishers who have decided to not engage with the ID-based proposals for the cookie-less future have adopted the popular alternative of contextual-based advertising and chose to reorganize their sites and their business models and targeting of individuals based on the contextual elements of the content that they create and publish.

I would warn though that if we're going to go in this route, just prioritize what is valuable about the media and not continue to optimize certain aspects of it and not develop taxonomies where we can optimize certain content or further partition parts of publisher content. But yeah, I think that's the ideal scenario going forward.

Nazanin Andalibi:

Piotr or Dan, do you want to come in?

Dan Calacci:

Yeah, sure. So, I really view, at this point, in time in 2022, data collection at work, in the consumer context, this blows of data or really the substrate of our current political economy. How data gets used, what systems optimized for using that data fundamentally change the economics of how platforms work, the kinds of content that people engage with and discuss political conversations. And so, if you do it from this kind of scale, it really becomes far less about we've been talking about privacy or just rights to data and far more about the actual systems that data flows and the algorithms that get developed, support, and facilitate. And so, I really agree with Nazanin about examining harm and looking at how

processes inflict harm, but also fundamentally because the goal of so much data collection and use is optimization or algorithmic development.

There are two pieces here. One is when those algorithms work, and one is when they don't. And when these algorithms don't work, when they inflict unintended harms, we sort of have a lot of scholarship around how that should be handled in part around disparate impacts, around racism, and bias. But there are all sorts of cases where emotionally AI works well and inflicts these psychological harms on workers in particular contexts that have contextual meaning because it serves an agency, it makes them feel surveilled.

And so being able to have regulation that flexibly understands or can adjust to contextual harms quickly and easily as they develop in the future, I think is really, really crucial and offers people a way to have some agency in what these systems are optimizing for. That sort of a call for governance of these technologies, not just of the data, but also of the algorithms that get developed, which is kind of a big call. I don't think that anyone knows exactly how to make that work or happen yet, but at the very least, we can start with trying to look at contextual harms in the workplace, in particular, consumer context, whether it's advertising, platform work, other low age work, white collar work, and trying to identify them with the people who experience those harms and working from there.

Piotr Sapiezynski:

I would just add that the framing of transparency, privacy, and consent bias. These approaches center technology work from the assumption that this technology will happen and now we're just trying to make this technology better. And I think what's really missing is the ability to question whether this technology should exist and operate in the first place. And on another note, I also wanted to add that even privacy is often weaponized by platforms to counteract transparency.

So, we have had a lot of pushback from Facebook on transparency saying that they can't really be transparent because it would come at a privacy risk to their users. And I think it's very important when we hear this to question whose privacy is really being protected here? And in the case of Facebook, this was the privacy of the advertisers, not actual users as we think of Facebook users. So, I think we really have to be careful about not allowing powerful actors to subvert these framings also against the kind of goals that we're trying to achieve here.

Amba Kak:

Okay. I'm tempted then to just end because that was such an excellent provocation. I just wanted to say a big thank you to our presenters today. Most of their papers are available on the Privacy Con website, and so I encourage everyone to definitely check them out. And thank you to our audience. I hope you enjoyed this discussion. We certainly did. And stay tuned for the next session. Thanks, everyone.

Sarah Myers West:

Hello everybody, and welcome to our second panel here at Privacy Con. My name is Sarah Myers West and I'm a Senior Advisor on AI at the Federal Trade Commission. In this session, we're going to be talking about automated decision-making systems. This session is structured a little bit differently in a conversation format, and we have a really fantastic group of experts who are going to be here in dialogue with one another.

Just to set the scene, we're here to talk about automated decision-making systems, which are technological decision-making systems that can be used in both public and private sector contexts to make determinations that affect people's lives in hiring, in housing, education, criminal justice, and elsewhere. In preparation for this session I spent time looking at different work about automated

decision-making systems, and one of my favorites is a paper by Rashida Richardson, who is an assistant professor of law and political science at Northeastern University.

She describes ADS as systems, software, and processes that use computation to aid or replace decisions, judgments, or policy implementation that can impact opportunities, access, rights, and judgements or rights and/or safety. These systems can involve predicting, classifying, optimizing, identifying, and recommending. For full disclosure, Professor Richardson is currently on leave and is at the FTC.

I think the critical point here is that these are computational processes that are often afforded a wide berth because they imply a kind of technological objectivity that may not always hold up under closed scrutiny, and this kind of scrutiny is badly needed given their effects on consumers' lives.

We're fortunate today to have a group with deep expertise in doing exactly that. So, I'd like to start by introducing our three panelists. We have Arvind Narayanan, who is a professor of computer science at Princeton. He co-authored a textbook on Fairness and Machine Learning, which is available online, and he's currently working on a book titled AI Snake Oil that I hope we'll get to discuss in more detail soon. He is a recipient of the Presidential Early Career Award for scientists and engineers, and he's thrice the recipient of the Privacy Papers for Policy Makers Award.

Second, we have Deb Raji, who is a Mozilla fellow and a computer science PhD student at the University of California Berkeley. She is interested in questions on algorithmic auditing and evaluation. Deb has worked closely with the Algorithmic Justice League to highlight bias in deployed AI products. She has worked with Google's ethical AI team and has been a research fellow at the AI Now Institute and that partnership on AI. She was recently named to Forbes 30 under 30 and MIT Tech Reviews 35 under 35 innovators.

Last but not least, we have Michael Veale, who is associate professor in digital rights and regulation at the faculty of laws at University College London. His research combines law, computer science and human computer interaction. It covers areas including machine learning, platform and protocol governance, and the governance of privacy enhancing technologies. And his work has been cited and applied by regulators and policymakers around the world. Dr. Veale sits on advisory boards for digital rights organizations, including Foxglove and the Open Rights Group, and the UK data protection regulator at the Information Commissioner's office.

So, to kick off the conversation, I'd like to maybe have each of you speak to a broad question, which is what do you think is most important for technologists and policymakers to understand about automated decision-making systems and accountability? And maybe we can go in order and start with Dr. Narayanan.

Arvind Narayanan:

Great, thank you so much. It's great to be here. My view on this is that accountability for flawed, algorithmic decision-making systems doesn't really exist today. I know that's a strong claim. Let me try to defend that and explain what I mean. Algorithm audits today are done by a ragtag group of people, a few investigative journalists, some nonprofits, and some academics. The problem is we're outnumbered something like a hundred to one by the companies that are out there putting flawed algorithmic decision-making systems on a regular basis. Most of those never get audited, particularly egregious cases might get audited. The Gender Shades study is a good example, but that's not the norm. There's a reason we still keep citing Pro Publica's study of COMPAS from way back in 2016. I mean, it was a good audit, but it still remains the best example we have of an algorithmic audit because there's no systematic program anywhere to repeat this on a regular basis with different systems and different geographic areas, even within a single domain like criminal justice.

Forget about the dozens of other domains in which these systems are employed. And that is not the only barrier. There is also the lack of access to these systems. There is only so much that auditors can do if they are third parties without any privileged access. Now, what audits are good for... They're not useless. What they're good for is alerting us to the fact that there is a general problem with discriminatory AI or flawed AI. What they're not good for is as a way of going after specific instances of malfeasance in a systematic way. And I'm sorry to say this, but I think the FTC and other regulators are also similarly vastly outnumbered by the companies. So, I think there is a big gap right now. If we're serious about closing that gap, perhaps we need to look at a different model like anti-regulation or private right of action. I don't know what the right answer is.

I'll leave that to legal experts. But I'll point out that I think what we have right now is not working. And the last thing I'll say is that we saw the same thing with privacy. Over the last 20 years, privacy research has been very good at alerting us to the fact that we're constantly being sped on, but that research by itself wasn't very effective at getting companies to change their behavior because the chance that any specific company... Well, except for Facebook. The chance that any specific company's privacy violations would end up in the news was very small, so companies pretty much kept doing what they were doing. So that's my view. I'm particularly curious what Deb thinks about all this since she has done more audits than any of us here.

Sarah Myers West:

I'll throw it over to you, Deb.

Deb Raji:

Yeah, I was just reacting to that. I think that's a super interesting take. I agree that there's no accountability in the market today, and that is why I'm very excited for Arvind's book where I do agree that there's a lot of snake oil in the market. There are a lot of things that are out there being sold for various reasons as AI products and causing havoc pretty much without any degree of accountability or in any systematic way. And I also agree that right now in terms of who is taking up the responsibility of doing auditing, you have a very sort of diverse group of people, especially in terms of external auditors, people outside of the institutions, the audit targets. It's a very diverse group of people doing it with various objectives, degrees of commitment to the accountability purpose.

If you're hired by the company, you might have a different commitment to accountability than someone that is coming in as an investigative journalist. So, because there are all of these different stakeholders with different motives, and different skill sets. The space is just incredibly... It feels a little chaotic in terms of who is participating currently and what standards of conduct they are held to or held against.

I will push back a little bit on this idea of audits, not necessarily us needing to think through alternative mechanisms. If you look at other industries, the medical device spaces one I really like because I think that they've thought through how to leverage audits pretty systematically, where in order for your product to even qualify for entry out to the market, you have to go through an internal audit and demonstrate a particular amount of due diligence in terms of your evaluation process. And you have that reviewed by a regulator.

Then you also have a group of audit organizations that are appointed by the regulator. And the audit organizations also have certain privileges such as access privileges and privileges in terms of enforcement. And that just kind of solidifies their entire audit ecosystem and matures it in a way that we don't see for the algorithmic context. So, I think that there are examples and it's not just the medical device space. If you think of transportation and finance, there are industries in which the audit ecosystems have matured to become meaningful forms of accountability. So, I think that there is a lot

we can learn from those spaces as to how we can systemize audits and systemize the sort of necessary scrutiny that these systems have to go under before they even enter the market in the first place, especially in a high-stakes context. You were mentioning criminal justice, but there are a lot of AI machine learning tools in the medical space and hiring and all of these areas that we do care about.

I think that it's possible for us to... For each of these domains, especially these high-stakes domains, we can think through how to systemize the scrutiny of these systems before they enter the market, but even after they enter the market if you think of the FDA's post-surveillance mechanisms where they have incident databases and they have ways of making it possible for folks to report incidents and then have that kind of connected to a group of auditing organizations that can then investigate these further. So, there are mechanisms and there are strategies available from these other industries that I think the AI machine learning space can learn a lot from, especially as AI machine learning products enter these other industries. That being said, I do think that there are things about the conversation that always has been a little bit confused or muddled, and that makes it difficult for us to have clear conversations about what it even means to audit these systems.

So, for example, Sarah, you gave a really great definition of ADS and that is from Rashida's paper, which I love, and I cite all the time. And I think when I have conversations with people that have been doing algorithmic audits, a lot of them actually don't audit ADS systems. I think me, Joy, and a couple other people, we came in sort of auditing these APIs and also thinking through audits of automated decision systems. But a lot of the traditional conversation around algorithm audits were the audits of online platforms, and that is because that work came from the HCI community or the computational social science ESS community in the academic space. And so, a lot of the strategies that even we use are of derived from certain assumptions around the interactions between the users of these systems and the nature of these platforms, which is not the case for ADS systems. A lot of ADS systems, if you think of a hiring tool, the user of the hiring tool is the hiring manager, and it affects the applicant. So, the impacted population is sort of separate from the actual user of the tool. And that's very different from auditing Facebook's newsfeed where I am the user of Facebook, and so there is opportunity for me to collect my own data and donate that to an auditor. Whereas if I'm impacted by ADS used in a hiring context, I might not even know that the ADS was used at all. So, there is a lot of different dimensions to the strategy. There are a lot of ways in which the audit process itself and the audit design, the design of the methodology around auditing needs to evolve to this ADS context. So, for example, something I'm trying to advocate a lot for now is just the visibility of these systems.

It's very possible to be affected by many ADS systems throughout your day even, and not have any level or layer of visibility as to like who is using what tools and how it's impacting your life. This is as just a regular individual in society, but also as regulators, there should be some level of awareness and communication around which ADS is being used by which parties and which institutional users. And then also, another point that Arvind had kind of alluded to was access and how difficult that can be and intimidating it can be as an auditor to try to access these systems when again, you have so many different stakeholders between you and the actual... Forget even the vendor. You have the vendor, you have the institutional user, you have so many different layers between you and the actual system itself, so it becomes incredibly difficult to obtain the information necessary to do the audit.

And I could go on and on about all the different challenges. There's independence. We have papers on this. I definitely agree with Arvind that there are a lot of issues, but I'm also hopeful in the sense that I think that there are other industries that have addressed these issues in a very systematic way, and we can learn from those other industries. And we also should be very clear about the fact that ADS is a different sort of beast than online platforms, or there's differences that we need to be aware of when we think of the design of audits and the institutional design infrastructure, we would need to support

audits in this space, versus the online platform space, which feels like a different conversation actually in many cases. So, I just wanted to highlight those two points and then I'll join the conversation later on.

Sarah Myers West:

There's a lot that we'll, I think, circle back in and unpack. But first I want to go to Michael and ask the same question which is, what do you think is important for technologists and policy makers to understand about ADS and accountability?

Michael Veale:

Thanks, Sarah. I think one thing that's important to understand is that digital infrastructures go pretty deep down, and understanding the nature of the infrastructures, social and technical around automated decision systems and similar systems is really key and important to pay attention to. We have technical infrastructures, data linkage, sharing tools, cleaning tools, APIs. These seem quite boring, but actually they're really important parts, not just because they contribute to the outputs of these systems and the way these systems can fail or succeed in practice, but they also contribute towards the stickiness of these systems, they contribute to who makes decisions around them, the kind of decisions they can make, they can't make, and the ways these systems develop, decay, or improve over time.

There're also important social practices around these systems. I've done several studies talking to practitioners in different contexts around automated decision systems that they use. And you really see, particularly in the public sector context, that some systems develop with quite a lot of care. They developed often in-house by public sector organizations, and those individuals who are responsible, they might know the limits of these systems, they might be able to fit them into social routines in ways that they envisage, in ways that are useful. Those people might leave. Those systems might be moved into other contexts that do not have that social scaffolding around. People don't understand the limits or the roles that they were initially designed to play, and they start to become something very, very different, and they start to be relied on in ways that are totally unreliable, and they move outside of their original design, or just the ways that they fitted into people's understanding of what they actually were doing.

So, understanding what these systems do when they move and migrate over time, is really key. And all of these things, they draw attention to the way that, while the qualities of an algorithmic system are really important, we've talked about bias and other panels, not just as a way of the past. I've mentioned this topic as well over and over.

There is also the important task for us to look at how they shift power. And I think the FTC is a great place to think about this, how the introduction of automated decision systems moves decision making, discretionary decision making, system design, policy choices, both at public and private organizations into different hands. So, we see public discretion, discretion that might be effectively reserved or preserved by statute. We see that shifting and moving into the hands of designers, and those designers might operate across multiple contexts, multiple jurisdictions.

Deb gave the example of hiring tools where you start to get the same decisions that are made. And yes, the hiring manager might be using these tools, but they're relying on something that might have really challenging both systemic issues, but this company that's making this, is also designing it in such a way so that their business can expand, so their business can maybe take over more and more of the functionality or the function of that manager was doing before, take on more contracts, exist to be durable over time. Same with education, technology, moving pedagogical decisions away from teachers, moving them into the hands of other kinds of actors.

And we have to ask, I think where we want decisions to be in the 21st century, decisions around system design, decisions around how our social services and other important services function. On top of that, these systems impose a very particular problem frame. Really, there's the old adage, "Everything looks like a nail if you're holding a hammer." Well, this is really writ large in algorithmic decision systems, because they have to recoup costs by scaling. And that scaling implies really that you see this problem multiple times. You start replicating and seeing it in multiple contexts, where it may not exist, and where in order to actually solve or manage complex problems, there may be a need to develop a unique problem frame specific to that context from the bottom up. Yet you're imposing it on some top-down tools, and that hits a conceptual, it clashes in a very conceptual way.

Add to that the way the people involved in procurement are not really the same people who are framing the problems that both public agencies and private actors might have. These actors lose the ability to frame and design the ways in which computers may usefully fit into these challenges. We also see procurement actors becoming arms of large platforms effectively. We see them becoming licensed managers in the words of recent work by Tobias Fiebig and Seda Gurses, and others who look at how schools and universities move to the cloud. The actors, it used to be IT managers and they used to be able to customize and tinker a technology to make it work in level contexts, and now doing office licensing and other kinds of licensing for different companies' products, and they are becoming agents of these companies inside. So, we have to think of all these issues in the round, because if we don't, I think we're going to just be looking at the surface level and we need to look a little bit deeper to identify the rooted problems. Thanks.

Sarah Myers West:

Thank you. I think across each of your provocations, we could bucket two different kinds of categories of challenges here. One is a set of harms that are directly caused by ADS themselves: how they're designed, how they shift power, how they move in the world. And then another category that's come up are harms that are caused that when they can't do what a vendor claims that they can do, and the methods that we might use to seek accountability there. And so, I'd like to ask one more broad question which is, how should we be thinking about these two categories similarly as distinct? Do they require different kinds of accountability? I'll open that one up to the group, but I can call on folks too if need be.

Deb Raji:

Sorry, to clarify, you're saying the distinction between what Michael was mentioning around some of these supply chain, the way in which the system itself is leveraged by different actors, and the decisions are fed in by various stakeholders versus just the quality of the actual prediction itself.

Sarah Myers West:

I think, that's one subcategory under the broader category of harms that are caused by the automated decision-making system working as claimed, and then another category of when they're not capable of working as they-

Deb Raji:

Okay. Yeah. That's something that we talked a lot about following the facial recognition work, where facial recognition is definitely harmful when it works, and it's also harmful when it doesn't work, and it's actually a different set of harms when it doesn't work. People are at risk of misidentification and that can have catastrophic consequences with Robert Williams and the false arrest. But when it does work, that can also have catastrophic implications such as ICE's use of facial recognition to identify folks and

challenge people's rights to privacy and autonomy in society. So, I think something that I learned, especially because a lot of algorithmic audit work is around engineering responsibilities, so a lot of algorithmic audit work is really around accountability of those making decisions about the system, and gauging or judging, attempting to judge the quality of those decisions as it relates to the consequences that we're seeing on the other side.

And as a result, a lot of the focus does tend to be on that second bucket of if people are making certain claims about how well this system is performing, how can we assess these claims and push back on these claims? And I think there's advantages to operating in that realm in the sense that you're just analyzing claims being made by these vendors, and thus just challenging or providing a counter narrative to those claims. And so, there's a precision to the conversation that I think is difficult sometimes when making arguments around the appropriateness of the use of a particular system, where in that case, then it becomes a much more, I guess, normative argument of saying that the downstream implications of this has this negative impact and you're trying to convince an entity that, and the retort is always, but it works. So, it becomes a little bit of a challenge in terms of actually making the case of the use case, for example, being inappropriate, which is something that was very difficult on the facial recognition front.

In my opinion, I think you need both. So, for facial recognition, what ended up happening was, by pointing out how these systems did not work for particular subgroup, it opened up the conversation around how inappropriate it was to use in specific contexts. So, for example, a lot of the companies that we audited ended up pulling their products out of direct sale to law enforcement. And that was because by pointing out how these systems failed, especially for the marginalized folks that were disproportionately targeted by these systems, it opened up a conversation as to why are we using it in such a high-stake scenario such as law enforcement when this technology is inherently fragile and inherently limited and inherently falls short of the claims even being made about them. And I think that it was an interesting connection of those two paradigms where people are leveraging this technology when it works in order to, they're weaponizing it against particular marginalized groups.

And we were aware of that, and we were arguing for that, but it actually strengthened our argument to also demonstrate the ways in which the vendors' claims fell short. And that just highlighted how much more inappropriate it was for them to be using it in that high stakes context. So, I think that was an interesting outcome of my participation in the audit world was seeing how those two conversations were strangely connected or could feed into each other in interesting ways. I think something else that has been really fascinating when discussing the use case scenario, because people talk about that a lot with facial recognition as well, there's a whole camp in terms of facial recognition policy that's very interested in the idea of restricting the technology or banning it, but only on a use case basis. So, you can use it for this, but you can't use it for this type of policy strategy.

And I think it's always been very fascinating because people will claim positive use cases. So, they'll say, "Oh, we're using it to find missing children", for example. And then when it comes down to it, that positive use case is maybe a very, very small fraction of what it's actually used for versus what the main use case might be. And that was a very fascinating experience for me as well, where people often present this dual use analogy of, oh, you have this technology and it can be used for good or it could be used for not-so-great things, and it's depending on the user. A lot of vendors will try to push the responsibility onto a downstream user, and they'll say, "Well, our policy says that you should set the threshold to this, and our policy says that you should not use it for this horrible application, and if our users do that, that's out of our hands."

But I found it fascinating where you actually have a sense of what users are leveraging this technology for, and if it's disproportionately this negative thing and not this positive thing, then this conversation around dual use becomes a little bit skewed where it's like, "Oh, actually, if people are

disproportionately using your technology for a use case with negative outcomes, especially for marginalized groups, then we should be rethinking what it means to make that technology widely accessible." And I think that, that's something else that I learned from the facial recognition conversation where although you have technologies that could potentially, where you could brainstorm this positive use case, if it's disproportionately used for cases that affected populations have huge and very valid objections to, then that should actually factor into how available the technology should be and to whom that technology should be available to.

Sarah Myers West:

Michael, in your work, I think you've often made the point that ADS and AI systems are not deployed in silos, but that they're the product of business decisions and they're going to inevitably be shaped by complex business models. So, I wonder if you could tackle this a little bit in how these issues are showing up in some of the conversations you're tracking around a potential AI Act in the EU and other related regulatory frameworks there.

Michael Veale:

Yeah, sure. I have my view, and the AI Act view, which are vastly different. My view on this, building on what we've talked about already, there's really no need for these systems to actually function. A lot of the business decisions around these are around trying to platformize many, many other types of functionality that we see in the world. There are companies that want to sell, automate the decision systems, and they want to sell them for their own sake, and they want their product to be judged on its own merits and whatever sell of it's really, really good. There are more companies that are wanting to use this as a way into get many other business opportunities within computing, say around facial recognition, but also around health data. They also don't want to be involved in data cleaning. They want to be involved in provision of other services. They want to build on this and build other kind of verticals inside public and private institutions.

So, I think we have to see this as much of the groundwork as well, but for the business models, the groundwork that companies are setting up, say health agencies, health institutions to do for data linkage, they're promising magical algorithm decision making systems that may never exist and may never work, but in the process they're saying, "Before we do that, we have to get into data cleaning. Before we do that, we have to get into linking all your data together. And before we do that, we have to do this using a proprietary tool that you won't be able to move away from, because we are..." So, there's a lot of this light at the end of the tunnel in ADS systems that actually has a lot of reshaping.

Now, that's where the AI Act in Europe misses this out. So, the AI Act in Europe sounds wonderful and new and fancy, but ultimately it's the European Commission digging back to a regulatory mode they developed in the 1970s for regulating products. It's what we use in the EU. Obviously, we are in the UK now, but it's what is in the EU used to regulate PPE. So, masks, it's used to regulate high pressure canisters, it's used to regulate generally dangerous equipment. And the way it works, these have technical standards that get applied to them. These technical standards, we don't want legislators to be making these, they don't really know what cloth masks should be made out of, so it's outsourced to private standardization bodies. These are bodies that fall under the umbrella of the ISO, but actually within that set satellite, which some are European subset of the ISO or private bodies still. Now, that's all well and good if you're looking at standardizing what a mask needs to be made of in order, it's written not to let through certain kinds of particles.

But what we're seeing in the AI Act is the delegation of decisions around fundamental rights, delegation decisions around bias, transparency, human oversight, cybersecurity, things that actually have quite

value laid dimensions, decisions about how much risk we're willing to take on in certain high-risk applications. The commission delegating those decisions to private standards bodies which are not very inclusive, heavily industry captured and produce documents which are behind copyright walls anyway. You have to pay about $150 in order to read them. So, this is a first challenge. The second challenge we see here is just a category error. This law sees AI as a product, and for the reasons I think we've talked about already, it's not just a product, that's one view of it to see an AI system as a product. It's not even a service, it's a way of reshaping power relations between organizations.

It's a business model, it's a supply chain. All of these ways of looking at AI show different in automated decision systems in general, show different approaches to regulating it. So, I think we need to move away from the easy view, which is to see something as a product, just another risky product we want to regulate that we know how to do, and say, okay, through some light it might look like this, but not through all lights, not through all problems. And the problem with the European Commission's AI Act proposal, is it actually looked at through that light to the exclusion of all others. It's like a preemption instrument in the US, so it will preempt member states from acting in another way, preempt it from seeing ADS systems. ADS is in a different light, and that's what I'm pretty worried about. And it misses out a lot of these complex supply chain and other issues, which I think we'll probably talk about a bit more.

Sarah Myers West:

Arvind, I'd love to circle back to you, because I think one thing that your work across a number of fronts has done, is to document that there are really significant reproducibility failures in the research around AI, around machine learning, that often there's an underlying over optimism about whether systems can work as claimed. And when others try to recreate published research, they're not able to for a variety of reasons. So as another vantage point on these issues, I'd be curious to ask, what do you see as some of the driving factors underlying this problem in AI research, and what does that mean for ADS that are being applied in commercial context?

Arvind Narayanan:

Sure, thank you. This is work with my collaborator, Sayash Kapoor, and we've been compiling evidence of what we think is an ongoing and actually exploding reproducibility crisis in applied machine learning. We're talking about machine learning applied to fields like healthcare or political science or what have you. And if you search for something like Princeton machine learning reproducibility, you'll find what we've put out there. What we've found is that in many fields, when people systematically look at a body of applied machine learning research, and they try to see if those papers meet basic reproducibility criteria, the majority of them tend to fail. Some of these examples are really egregious. For instance, there have been, I think, thousands of studies of AI for COVID, and almost none of them have panned out. And a lot of it is because of things like testing on a dataset where all the training and testing on a dataset where all the positive examples of COVID-19 were adults, and all the negative examples were children.

So, you end up building an age classifier instead of something that can look at chest radiographs and try to detect signs of illness. So that as an elementary blender, but other examples are much more subtle. And the reason performance evaluation of machine learning is so subtle, it's a little bit of a dark art, is that what we're doing in machine learning is we're claiming that we're going to be able to predict how a classifier will perform when deployed in a particular context, by only looking at performance on some previously collected past data set, so that's the big disconnect. And there's essentially an infinite number of possible differences between our deployment scenario where we want to use it, and our past

observational data. And unless we have thought about and corrected every single one of those possibilities, our performance evaluation is probably going to be wrong. And the tragic thing is that because of some subtle technical reasons, we almost always end up overestimating performance, not underestimating it, so we end up being over optimistic.

So, I think that's what has led to the reproducibility crisis in applied machine learning research. But let's back up. What does any of this mean for ADS? And so, I think there are three big barriers to accountability and ADS, we've been already talking about some of them. The first one is that companies often don't even make their performance evaluations public. They might simply put out a press release saying that their system is 90% accurate, whatever that means. That's the first barrier.

But even if they do make their evaluation public, it might not be peer reviewed, and so it's questionable how much confidence we can have in it. And the third barrier, and this is where it gets back to what I was talking about, is the third barrier is that even if something is peer reviewed, I question how confident we can really be in that published number, because as I've said, even peer review has not proved to be an effective way of having confidence in the performance of machine learning models. So, I think yeah, we need something deeper than what we have now in terms of knowing how well a system performs, and having confidence in that assessment.

Deb Raji:

Oh, I don't know if I can step in.

Sarah Myers West:

Sure.

Deb Raji:

Yeah. I was going to say, I definitely resonate with those three points. I was going to say, in terms of the peer review point, especially around the level of oversight and scrutiny people have over the claims being made by these companies, and this also reflects Michael's point around who sets the standards and who is actually in the room when some of these decisions are made as to what the expectations are around these systems, I find overwhelmingly, you have the companies articulating which benchmarks they would like to be evaluated on, what metrics that they would want to be evaluated on. And I think that, that's why third-party auditing has been really interesting to watch, just it flourishing as a community, because you have a diverse range of individuals and stakeholders representing very different groups.

And these different groups are worried about different things. And so, they come up with their own benchmarks and they come up with their own metrics. And as a result, you have these interesting counter narratives to the vendor's narrative about how well their system works. And I think that, that's going to be this really interesting and critical point around accountability for these systems is, what does it mean to move beyond just each vendor communicating about their technology and whatever the way they want to communicate about their technology. What does it mean to actually require some different perspective factoring into how they assess their systems and how they communicate about the performance of their systems.

I think that point is going to be incredibly critical to actual accountability in this space. And that could mean involving a broader range of stakeholders in the standards making process, and then requiring these companies to adhere or communicate performance according to those standards, but it could also just involve opening up access and allowing for this oversight to happen, allowing third party auditors to be able to come in and access the information they need to tell their stories about how well the system

is working, push back against the vendor's narrative about how well the system's working. There's a bunch of different ways to approach it, but I think that, that's a very critical point. Just wanted to plus one that.

Sarah Myers West:

Deb, just as a follow-up question, I know in some of your research you've looked at the landscape of the different types of actors who are showing up in the auditing space. I wonder if you could speak a little bit to that. Who are the range of auditors, and what are some of the differences between them and their incentive structures, how audits might differ depending on who's doing the auditing?

Deb Raji:

Yeah, for sure. And this was a project that I did with the Algorithmic Justice League, and it was looking at auditing the auditors. We did a survey of about 150 folks in the algorithmic audit space, and try to understand who was in that space and what were they doing and what were their motivations and their goals. I think there are things that unify those seeking to use auditing as a mechanism of accountability in the sense that everybody, at least at a high or normative level agrees on the goal of accountability, whether it's internal or external accountability. They all want to be able to judge the quality of these systems and look at things through the perspective of how well are the decisions being made about these systems impacting the population that I'm meant to represent. I think that the positionality of these auditors vary widely.

So, you can have a group of folks that we've classified as internal auditors where they have some kind of contractual relationship with the audit target. And that could mean as employees of the company on an internal audit team, interfacing with leadership in order to articulate internal expectations for performance, and then gauging how well the entire system that produces the output product correlates with that internal expectation. And these are typically in modern tech companies, this would be the responsible innovation team or the ethical AI team at the big tech companies. And even at smaller startups, you usually have a compliance team or perhaps even an explicitly internal audit team that goes around and looks at how the different components, not just in engineering, but maybe even in product or business or legal, connects to these internal expectations of performance.

And sometimes also, companies will hire in consultants who have their own proprietary standards. ORCAA has their own standard for performance, and they'll bring in these consultants, sign a contract with the consultant, the consultant comes in and they assess the system. And there's a lot of pros to having internal audits. There's other industries, like I mentioned, more mature audit ecosystems and finance, the medical device space and transportation, especially in aerospace in particular, they have internal audit regimes. It's a long-standing practice that's happened in other industries in other spaces successfully. But there's requirements for that to work out, for one, you need the buy-in of leadership. Leadership has to actually take in the recommendations of the internal auditors and implement them in a way that's meaningful, and that doesn't always happen and that's one of the challenges they face. But on the other hand, they can look at the system pre deployment, they have full access to the information they need. So, they tend to actually be very thorough and effective in terms of articulating really clear views of what's happening internally with the system.

And they can also intervene before the system reaches the market and hurts anybody. The challenges though, which is exactly why we have external auditing, is that a lot of internal audits are not published for reasons that make sense, like liability concerns of the company. And so, you have these external auditors, and anyone outside of the system really that doesn't have any visibility into what these internal auditors are actually discovering and the problems that they're finding. And that's why I really

like... There's mechanisms that people are, like the ICO, the Information Commissioner's Office in the UK has actually a mechanism where they say, "Okay, well, you have to produce these internal audits and then also report it to a regulator." And that can be an interface between the company and potentially some external parties that have an interest in that report. So, there's mechanisms and discussions around how we can improve the limitations of these internal audits, but I think that, that's one category of audit that's definitely growing very rapidly.

The second category is what we call external audits. And these tend to be third party organizations or individuals, and they are looking at the system from the outside. They don't have any contractual relationship with the system. A lot of their audits are complaint initiated, which is why I mentioned the importance of incident databases where they will hear about an incident, something happening, the algorithm going wrong in some way, and that will prompt them to investigate the system via an audit. But since they're on the outside, they typically don't have any level of access to these systems. In fact, there's been already a couple cases with Facebook retaliating against external auditors leveraging privacy and anti-hacking laws in order to push back against the attempt of these auditors and watchdogs to access the information they need to evaluate these systems externally.

So, it's really difficult. Access is a huge challenge, but also even identifying the target of the audit is a huge challenge. For example, if you get a complaint that someone suspects that there might be an algorithm involved in the process through which they applied for housing and got rejected, it's really difficult to figure out what vendor the landlord was using. Even though you have a suspicion that there's an algorithm involved, it's incredibly difficult to actually identify the vendor, to even isolate the model to be able to properly audit that system. So, there's a lot of design challenges to external audits that make it incredibly difficult to execute. But on the other hand, they are completely independent of the audit target. So, there's a lot of meaningful work that's happened throughout investigative journalism, civil society. There's now specialized law firms looking at external audits, and of course, academia and regulators, a wide range of different types of stakeholders playing a role in terms of examining these systems from the outside, and working through these limitations in order to get more meaningful reports of how well these systems are working.

So yeah, a very wide range of folks involved. I think one of the main things we learned from that study was that given the diversity of who's involved, there needs to be some level of accreditation or training involved as well, where the quality of the audits range drastically. It's not enough to just do an audit. Not all audits are made equal. And so, there was a lot of discussions that we had following that report, and especially examining the survey results of what does it actually mean to do this successfully, and what are the criteria for what constitutes a high-quality audit. So, for example, and this is something that the finance industry has faced, a lot of auditors, even external auditors have issues of independence and conflict of interest, and that shapes the nature of how they report their results and exactly what they're auditing, the scope of their audits.

And then also, sometimes there's a lack of technical capacity. And so, the audits don't say the full story of what's actually happening or they're not able to investigate things as thoroughly as they could. And then to Michael's point earlier, sometimes you're just isolating things to this specific model and you're just looking at a static data benchmark when maybe the issue is a broader, more systematic issue. So, there's so many questions as to what an effective audit procedure looks like, and how to enforce any level of auditor conduct standards and accreditation. If someone is auditing something for, an FTC investigation, what's the quality of the audit that's required? What's the credentials that the auditor will need? This is a huge conversation with the Digital Services Act right now, is just who actually qualifies to play this role and can be trusted to play this role, especially if they're given any privileges such as access as part of their role as an auditor.

So yeah, it was a really great experience to scope what was out there, but there's definitely a lot further, there's a lot more we could do as a community to solidify the group. I think, one final very small point I'll make is that something that's interesting about the audit community as well, is that everyone is doing their own thing. So, there's not as much communication as there could be, and there's a need for coordination and communication across the community and across the field. And the way that I discovered this, was working on the open-source audit tooling project with Mozilla, where people will build tools to execute an audit, and they won't necessarily share those tools with other auditors just because there's no communication and visibility as to what other people are doing.

So, you end up with multiple tools doing the exact same thing, built ad hoc by different auditors for their specific audit, for their specific target, their specific goals, and not a lot of shared resources and infrastructure. And that makes the whole process even much more difficult to execute. So, things like that are definitely plaguing the community in a serious way. And I think one of the conclusions of our paper was that there just needed to be more coordination and communication across the community as well, and that there's roles for different stakeholders in terms of pulling people together, and that could solve a lot of problems for us.

Sarah Myers West:

Thanks, Deb. I'd love to circle back to you Arvind, because I think in your introductory remarks you made a number of points around some of the limitations of auditing frameworks, and that your work has led you to be more in favor of ex ante approaches to regulation. I'd love if you could talk a little bit more about that and about your current work around AI snake oil and unpack that framework for us a little bit.

Arvind Narayanan:

Sure. Those are two different things. Maybe I'll touch briefly on the auditing point and then go to AI snake oil. I think what I was saying about ex ante regulation was very much in line with a lot of what Deb was saying, which is that there has to be some legal incentive for the audit to happen as a matter. Better, of course, especially before the product reaches the market and then ideally on a periodic basis.

And for that to happen, like, again, the example for medical device regulation, I think that's a really good model to look at, I think something like that would be helpful for algorithmic auditing.

Another thing that would help is ... I was pointing to the numbers problem. Just the number of, for instance, people in academia who are doing this is really minuscule, and I think there are various kinds of incentives that can shift those numbers, some of which are easier than others.

I'll talk about academia since that's what I'm most familiar with, but I'm sure for journalism, for civil society organizations, there are things that can help as well.

In academia, one thing that's very simple that can help is, let's say an entity like the FTC that finds an audit to be very helpful in your work gives that paper an award; this really increases the recognition for that paper. It costs nothing to provide an award, but it's a signal that this work is malleable for regulators.

And this sort of thing helps researchers who are doing this work justify it to their tenure committees and others, because ultimately, that is a constraint. If they're not going to get credit for it, the work is not going to get done.

So currently, auditing work faces a big barrier in publishability, because once you've shown how to do a particular kind of audit, doing it in a repeated way is not going to get you repeated publications.

So, if we want to better align the work coming out of academia towards things that are more actually helpful on the ground, we perhaps need ways of recognizing that work that go beyond getting publications out of it.

So that's one suggestion for how to perhaps start to overcome some of those limitations that I currently see in the auditing space.

So, moving to AI snake oil, I'll touch on this really briefly because I know we're getting close to the end of the hour here.

Let's talk about hiring tools for just a second. I know both Deb and Michael already touched on this.

Really this is the reason I originally started working in this area. I was shocked when I learned back in 2018 that there are hiring tools that claim to predict job performance, not even by looking at someone's resume, but by looking at a 30 second video of them speaking, and then not even based on the contents of the video, but by analyzing body language and facial expression and things like that.

And so, it struck me that these are really elaborate random number generators and really nothing more. There's no known way in which these can work. If anything, they're just picking up on biases and stereotypes in the way people speak and so on.

And yet I was shocked to find that millions of Americans have been put through these demeaning kind of job interviews. There still remains no evidence of their effectiveness. Unsurprisingly, I don't think they work at all.

So, what I want to say about this is that this is different from some of the other kinds of ADS failures we've been talking about today.

I think we should approach a tool like this with a high degree of skepticism. I don't think job performance is really predictable, especially based on this kind of video analysis.

And so, it's not so much a matter of, is a particular tool failing? It's more a matter of, what basis is there to believe that this is even a meaningful thing to tackle using an automated system?

So, I think for cases like that, for cases where we have no affirmative evidence and strong reasons to believe that you can't actually predict something, I think there needs to be a heightened standard of scrutiny.

I think the EU has a risk classification system. I wonder, in addition to the risk classification, whether this kind of classification, you know, what does the scientific evidence say about whether this is even a problem you can solve? I wonder if that is a useful legal classification for what kind of evidence we look for, who has the burden of proof and that sort of thing. I think companies should have affirmative evidence before they're allowed to put something like this in the market.

Sarah Myers West:

Michael, you look like you're about to jump in. Do you have any reactions or final thoughts?

Michael Veale:

Yeah, just ... I know that's the important part. I know there is a risk that these risk classifications sort of say, well, this system works by kind of giving it regulatory legitimation.

But I wanted to highlight one thing that hasn't come out yet, which is the power of auditors and actually the concern of audit as a competition problem.

As Arvin said, auditing isn't very sustainable for academics to do. It's unclear exactly what the business models of audits will be. But what is very clear is that auditing is something that may be captured by platforms.

Auditing may be a thing where you say, "Okay, actually, you want to buy a trustworthy system. Well, we bundle it in with APIs. We bundle auditing in, we bundle expertise. And in fact, you can't move to another platform or system because no one else could possibly audit it. Your supply chain relies so heavily on our upstream services.

For example, our large models we've trained that no one else has, we're the only ones that can audit that, and we can do a more holistic audit and your insurer is going to have to rely on that."

So, I think audit ... When we hear about audit, audit, audit, who are these auditors and what is the long-term business of auditing, and how do we make sure that auditing doesn't become another way that large companies can capture and dominate a market?

Sarah Myers West:

Thank you, Michael. We're out of time. I'd love to say thank you so much. This is a really rich and wide-ranging conversation, and really appreciate the expertise of all the folks on this panel.

We're now going to be going to break, and we'll return at 1:00 PM for our next session. Thank you all.

Genevieve Bonan:

I'm Genevieve Bonan. I'm an attorney in the Division of Privacy and Identity Protection of the FTC, and I have the pleasure of moderating this panel with researchers Kaiwen Sun [inaudible 02:26:09] and Noura Alomar [inaudible 02:26:10].

Just a reminder: the agenda, research papers and bios of our esteemed panelists are available on our website, and you can follow along on Twitter at FTC, #privacycon22.

Kaiwen [inaudible 02:26:24], please begin.

Kaiwen Sun:

All right. Thank you so much.

Hello, everyone. I'm really excited today on behalf of our research team to present our work on child safety in the smart home, parent's perceptions, needs and mitigation strategies.

Next, please.

Families adopt smart home technologies for different purposes, such as security, convenience, entertainment, energy saving. Children inevitably become a part of those smart home experiences. As you can see in those pictures, they're being around robot vacuums, they use smart logs, use smart speakers, and they'll be monitored by security cameras.

Next, please.

However, privacy and safety issues arise, as news and media reported.

For example, Alexa told a kid to do a dangerous challenge; hackers threatened children through cameras that are installed in a kids' bedroom; battery parts of smart buttons pose choking hazard threats; and babies encounter robot vacuum, and that signals physical safety risks.

As you can see, all the examples here, they're all happening in the home environment. Any parents who have gone through childproofing their home would know how important child safety is. So, we all want

children to grow up in a safe home environment, and yet those smart home technologies can pose those safety and privacy risks.

So our study looks into how parents perceive and manage children's safety in the smart home environment that actually involves three phases. And next, I will discuss each of the phases in detail.

In the first phase, before parents purchase any smart home technologies, most of them were vigilant in thinking about safety considerations on two aspects. One is how they can avoid physical harms, and the second is to minimize privacy and security risks.

Here's a quote from P9; "Generally I wouldn't put something with them that I felt was unsafe or would invite them to do things that were either physically dangerous or would kind of invite inappropriate experiences."

Another parent said, "I think that privacy and security and safety, and all of that rolled into one, is the biggest concern for me. I don't necessarily want anybody else looking at my child while they're sleeping."

So parents, they would avoid buying any devices with motors or that could generate heat, and they only pick trusted brands that rarely have any negative news.

Next slide, please.

During the second phase, once children started interacting and using those devices, parents had three safety reevaluations.

Parents realized there's the physical safety incidents, such as smart vacuums, run over the child's toes, and children were also exposed to unsuitable content by using smart speakers.

Like P12 said here; "With a smart speaker, you can ask all kinds of questions, and maybe if you mispronounce something, you get something else that's more adult related content."

And also, there's risk from accidental use and misuse. The quote in the middle was from P20. He allowed his nine-year-old son full smart home access because he doesn't want his son to miss out on the smart home experiences. But P20 worried that, "Let's say I'm in the shower and my son tells the smart water system to raise the temperature to 150 degrees. I could get burned by the water and that risk is there."

And another parent also expressed the concern with accidental misuse. P16 talks about they don't want kids to be silly and say, "Alexa, open the garage door."

Next, please.

So reacting to those physical safety risks in the second phase, parents really, really struggled to find ideal mitigation [inaudible 02:30:34] strategies.

Some of them used [inaudible 02:30:37] childproofing techniques, such as keep the smart home devices out of reach of children through blocking or elevation; some used parental control features, if available, such as filter content or set passwords.

Note that these are not necessarily child-specific parental controls, but rather general ones.

Parents were also selective about children's access to certain devices or functionalities. For example, smart speakers can only be used in the public spaces at home.

Parents also tried to come up with different types of rules and boundaries to teach children. For example, P22 said he would tell his children, "Hey, if you see that there's a camera in the room, we're not going to get undressed. Stay dressed." And yet, if there's somebody yells out through the camera and it scares them, the children should just go ahead and unplug it.

Next, please.

So beyond all these immediate safety considerations and mitigations, parents also talk about this evolving safety needs as children grow up, which is not well supported by the current smart home technologies.

Parents proactively anticipated potential digital safety risks and the need to teach children digital hygiene and good habits.

P12 here says, "We haven't had in-depth educational lessons yet on privacy and informational security. I mean, as they get more uncontrolled access, we'll have to have more of these conversations."

Parents also talk about the need to adapt parental rules to cater to children's independence and autonomy as they grow up.

Like P11 said, "If the kids' got the Alexa speakers in their room, like in a private setting, especially as they get older, they can start to ask inappropriate questions or adult questions. So these are the sort of concerns that I have as they get older. And if they get to have a more private use of them, it'll come and I'll have to put a lot of thoughts into that at that point."

And lastly, parents talk about the need to build a trusting relationship with their children and respect their children's autonomy when they grow up.

Like P7 said here, "I cannot control what they're going to download when they're 13. I have to let go and say I've done the best that I can."

So as I presented parents' perceptions and mitigation strategies, what are some of the key takeaways?

Next, please.

The first one is, parents are really concerned about children's physical safety, digital safety and privacy risks in the smart home context before they purchase the device, during the devices are being used, and when children are grown up. So it's a whole evolving process and parents really try to manage those safety risks and issues extensively.

Second is that smart home technologies, they're unique in that they amplify child physical safety, digital safety and privacy risks because they are physical objects situated in and modified in the home environment.

They are windows to the internet. They could collect a lot of data about children and their family. And they enable children's control via voice command or automations.

Next, please.

Despite of all these physical and digital risks, there's really a lack of support and resources. So parents become the sole gatekeeper of child safety.

In particular, there's a lack of child safety centered design features from those smart home products.

So think about it; if a washing machine has a child lock as a safety standard, why shouldn't a robot vacuum have a child lock? I know some of the brands are starting to have child locks, but it's not a standard yet.

There's also the lack of educational resources for parents to stay informed of how to keep their children safe.

A parent said here, "It will be better if those big smart companies or tech companies can have parental tutorials. I feel like if you're not tech savvy then it's hard for you to get on top of everything. It's easier on this age when children are younger, but when they grow more tech savvy and grow older, they'll get better at breaking parental control rules."

Next, please.

Parents also struggled with the lack of granular parental controls, because the current all or nothing smart home access control lacks the support for a more tailored children access and news.

But it's important to remember, we talked about earlier that parents want to include children in the smart home control experiences despite the lack of product support, and children are users of smart home technologies.

So I have a quote from P17 on the bottom left that says, "Let's say I want to give my son access to a smart thermostat. There could be an ability to add a child account that only allows the child to manipulate things by a few degrees or so."

And lastly, parents point out this lack of transparency about children's data collection and news. A quote from P14 here, it's a little long, on the bottom right side, but it really captures the essence here.

So, the parent said, "The parental settings that I found was not advertised to me. I had to go and Google parental control by myself. I know that Google knows I have young children, the amount of diaper ads I get. So, at no point did they go, 'Hey, here's how to protect your children', or 'Hey, here's how you can control your children's access.'.

As a parent, I would like to have more information on not just what the kids are doing on the smart speakers, but also what is known about my kids. I would like the transparency from Google to say that profiling children exists, and I would like the ability to delete it."

Next, please.

In summary, our study identified parents' perceptions and mitigation strategies of child safety risks in a smart home context. And we've also talked about how smart home amplifies child physical safety, digital safety and privacy risks while contrasting this to the lack of child safety centered and privacy protective smart home features, designs, controls and resources for families.

And finally, parents, they really tried to be the gatekeeper for child safety, but they struggled. They struggled with the limited supports and resources. So, this is where we're calling for companies to be more responsible, accountable and transparent with regulatory oversight.

Thank you so much for listening. We have our paper linked here.

I'll pass it on to the next speaker.

Genevieve Bonan:

Thank you so much, Kaiwen [inaudible 02:37:50].

Noura, please take it away.

Noura Alomar:

Thank you, Genevieve.

My name is Noura [inaudible 02:37:56] and I'm a PhD student from UC Berkeley. And today, I'll be talking about our work, which investigated the privacy compliance processes followed by developers of child directed apps.

This is joint work with my advisor, Dr. Serge Egelman [inaudible 02:38:09].

Next slide, please.

There are several reasons that motivated us to conduct this research, the first of which is that researchers continue to find privacy issues in various types of mobile applications, which include collecting and sharing users' personal data without having proper legal basis for doing so under applicable privacy laws.

We also observed that, while privacy regulations continue to be enacted in countries around the world, it's unclear how our developers are dealing with the legal requirements of laws that are applicable to them and what their compliance processes are.

And since the legal requirements surrounding the handling of data collected from child users are known to be more stringent than those that apply to general audiences, we believe that studying the compliance processes followed by developers of child directed apps presents an opportunity for understanding the challenges experienced by developers when trying to build compliant apps and also the kind of support that they can be provided with to help them understand their compliance obligations.

Next slide, please.

In this research, we focused on investigating the processes followed by developers to comply with three privacy laws that have [inaudible 02:39:15] provisions that apply to developers whose apps are available to users in the United States or Europe.

These laws are the Children's Online Privacy Protection Act, COPPA, which applies to children under the age of 13 in the US; the General Data Production Regulation, GDPR, which has provisions that protect users under the age of 16 in the EU; and the California Consumer Privacy Act, CCPA, which applies to California residents.

The CCPA requires obtaining verifiable parental consent before opting-in children under the age of 16 to the sale of their personal data, and explicit consent from those who are between 13 and 16 years of age before selling their data to third parties.

Next slide, please.

Under these laws, developers are required to be transparent to users by disclosing the types of data their apps collect, the kinds of third parties that receive users' data, and the purposes of their data collection and [inaudible 02:40:08] activities and their privacy policies, for example.

These laws also provide users with other rights, including the ability to withdraw their consent, inquire about how their data is used, or request a deletion of their data.

Next slide, please.

In our research, through a combination of technical app testing and qualitative analysis of self-reported responses from app developers, we were able to investigate whether developers understand their compliance obligations, whether they have processes for compliance and for vetting third party [inaudible 02:40:39] for inclusion in their apps, whether they're aware of their app's data collection and sharing practices, their consent [inaudible 02:40:46] procedures, their experiences with App Store policies, and the challenges they experience when trying to comply with applicable privacy laws and App Store policies.

Next slide, please.

For that, we deployed an initial survey to understand organizational privacy practices to developers who published child directed apps on the Google Play Store.

Responses to the survey were collected from April to August 2021. The results of this survey inspired us to design a second survey for personnel within the developer organizations to further understand their perceptions of compliance and what their organizations do to comply.

We deployed this survey from August to October 2021. However, the data we collected using the two surveys was still not sufficient for understanding compliance processes, and that's basically the reason why we also recruited app developers for semi-structured interviews from September to December 2021.

Our studies were not only based on qualitative data. We additionally analyzed developers' apps and compared the results of app testing with developers' responses to our surveys and interview questions to understand whether they actually comply with applicable privacy regulations.

In our interviews, we also presented developers with the results of our technical analysis of their apps, discussed our findings with them, and heard their thoughts on our findings.

By doing so, we were able to understand how they perceive their app's behaviors and the extent to which they believe that their apps are compliant with applicable privacy laws.

Next slide, please.

We also discussed with interview participants whether they're aware of the privacy compliance configurations offered by a number of third-party SDKs integrated in their own apps.

These configurations allow developers to ask SDKs to treat data received from child users in accordance with what applicable laws require. For example, to restrict the kind of processing performed on data received from child users.

This screenshot provides an example of server-side privacy configuration offered by a third-party SDK called Unity Ads, which should be configured in child directed ads that use Unity Ad's services.

For example, these configurations allow developers to request serving their child users with contextual ads instead of targeting them with behavioral ads.

Before conducting the interviews, we tested the configurations offered by a number of third party SDKs that are used in child directed apps. We integrated the SDKs in apps that we created ourselves and then we tested the integrations to observe the traffic that gets generated from the apps as a result of integrating the SDKs.

Our testing of these configurations helped us collect a list of signals that we were able to then use to understand whether real child directed apps are using the SDK compliance configurations correctly or not.

We identified cases of SDKs that were misconfigured in real apps and discussed that with the developers during our interviews.

Next slide, please.

After analyzing the data collected through the first survey, we found that 42% of the 50 developers who responded to our first survey indicated that they don't have formal processes for [inaudible 02:43:47] privacy issues during the software development life cycle.

We also found that only 36% of them reported that they obtain parental consent across all their apps prior to data collection.

Our survey and interview results interestingly also showed that developers lack awareness that parental consent needs to be verifiable.

We found developers who use simple knowledge-based questions or age gates as their consent mechanisms, where they can be bypassed by children, and therefore, their means to obtaining parental consent, did not appear to be aligned with what applicable laws require.

When we discussed that with the participants during our interviews, some explained that it's the responsibility of parents or App Stores to verify that consent is obtained from guardians and not children.

Others who responded to our surveys believed that their apps are not dangerous, and therefore, they didn't see the need for having verifiable parental consent screens in their apps.

For example, one interview participant mentioned, "I didn't know about the thing where [inaudible 02:44:39] make sure that it's the parent and not the kid."

Another survey respondent mentioned, "Our app's not dangerous when used by children, so there is no need for parental approval first."

Next slide, please.

We investigated developers' familiarity with four privacy laws in our second survey as well, which are COPPA, GDPR, CCPA and the California Privacy Rights Act, CPRA, which would replace the CCPA.

We found that developers were more familiar with COPPA and GDPR compared to CCPA and the CPRA. In the survey question that asked about familiarity with privacy laws, we added a fictitious law that doesn't exist, which we called [inaudible 02:45:16]. We added that just as a quality control.

In the survey results shown in this figure, however, we found that 23% of respondents claimed familiarity with this law that doesn't exist.

We interpreted this finding as showing the level of [inaudible 02:45:29] of developers and their compliance obligations, particularly those who are independent developers, those who are not based in the US or the EU where some of these laws apply, or those who are part of small organizations that don't have access to legal expertise, for example.

Next slide, please.

Findings from our second survey and the interviews also show that many developers rely on feedback received from the Google Play Store or [inaudible 02:45:57] applicable privacy laws, and also assume that their apps are in compliance with applicable laws if they're accepted for inclusion in the app stores.

As for developers' [inaudible 02:46:07] concern regarding the consequences of not complying with applicable privacy laws, 68% of the 77 who responded to our second survey believe that they're unlikely to be investigated by regulators, and 44% were concerned about the possibility of removal of their apps from the Google Play Store if privacy issues were to be found in their apps. This highlights the major role app stores play in privacy compliance.

As for third-party SDKs, we found that only 31% had processes for [inaudible 02:46:33] third-party SDKs for inclusion in their apps and that [inaudible 02:46:37] might lack expertise in how to do so.

We also found that the majority of interview participants did not know that they had to configure SDKs for privacy compliance, with some indicating that they trust that SDKs provided by prominent companies are legal to use and that their apps that integrate these SDKs and use the SDK's default configurations are compliant by default.

Others believed that they did not need to do so because their apps were accepted by the store, and therefore, they assumed that they're in compliance regardless.

Additionally, when you ask about how challenging is privacy clients for developers in our second survey, only 8% of the 77 developers who responded believe that it's not a challenge for them.

Next slide, please.

And so, one of the main takeaways from our mixed methods research is that, while the majority of developers have awareness that this is a regulated area, they don't fully understand what their compliance obligations are and rely on external feedback received from app stores or external auditing services, for example, to determine whether they are in compliance.

Our results also show that developers are feeling the burden of compliance and need usable tools that help them test their apps for compliance, understand the compliance obligations, and get advice on how to build compliant apps.

This guidance could be provided by app stores during the app submission process, by third party SDKs to make it easier for developers to understand how to configure them for compliance, or by our development tools or IDEs that developers use to build compliant apps.

For example, developers could be provided with up to date [inaudible 02:48:08] to the various SDK compliance settings that need to be configured on their apps during the app submission process they go through before submitting their apps to the stores.

And since there's prior work that showed that developers are more likely to keep using the default settings of third-party SDKs, we expect modifying the defaults of third-party SDKs to be privacy preserving to help developers in their compliance efforts.

Finally, since our results show that developers rely on app stores for advice on how to comply, we also believe that better compliance rates would be achieved once the app stores provide better enforcement of the requirements of applicable privacy laws.

Thank you for listening and I would be happy to take any questions about our study.

Genevieve Bonan:

Thank you, Noura. Thank you both so much for your presentations. I know I personally enjoyed reading your papers and I'm so thrilled that you could present today.

I do have some questions for you.

I will start with Kaiwen [inaudible 02:49:02].

Noura covered this in her presentation, but I'm curious, Kaiwen [inaudible 02:49:06], what was your motivation for the research on smart home devices?

Kaiwen Sun:

Absolutely. Thank you for the question.

I think it's a combination of seeing the trend rise in society, also reflecting on myself, as parents.

So, when I was doing this line of research, it was back in the middle of the pandemic, 2020, and I remember we had to all quarantine at home. And my kid, at that, time was two. She used a lot of technologies. And there were a lot of errors and issues as I see her using these.

And also, at the same time, I read news articles reporting all these child safety issues and privacy problems start to emerge; how devices are collecting data and how kids are having a lot of usability, privacy, security issues with the devices.

So that was where I started to think about, "Okay, how are parents managing this during the pandemic with kids and devices?", and whether there's research on this.

And I realized that a lot of work have looked into children's privacy, but not necessarily safety in the smart home context. There's work on kids and mobile app use or in game settings and all that.

So, I think smart home is rising because these devices, they're not necessarily ... You don't really think about children initially when you think about smart home, but if you take a second thought on this, they're for home use, so children are all over the place at home.

That's where I think we need to place the focus on this particular type of devices, because they're so at hands, so easy and convenient for children to get their hands on. They're creative users, sometimes

they're rule breakers and all that. That's why we need to think about how these devices can be designed and developed with children's unique needs in mind.

Genevieve Bonan:

Sure.

Can you expand on that a little bit more? You said that the parents are the gatekeepers for their kids' privacy and safety with the home devices. How can we help others be the gatekeepers of children's safety and privacy as well?

Kaiwen Sun:

That's an excellent point.

I feel like one thing is to think from the device perspective. Companies, whoever are designing those, they need to think about child users and the use cases; how these things might be used by children from different age groups and how children might misuse those.

I mentioned an example where there's misuse. Let's say a child user is trying to ask Alexa to automate or control different things. There should be some layer of gate keeping on that regard.

So, definitely, think about how can we keep children safe as a product design default, rather than, "Okay, let's just have kids' safety as a tagalong, rather than considering children as the key user groups as well."

And also, there's the support for parents. There's no place parents can find support, even though they tried.

So, I feel like big companies, like for example, Google, have internet safety related resource, but that's for content and internet use in general.

So, there could be more resources for families from the company side when they're designing those products; how they're thinking about providing resource or tutorials as the products are being developed, being iterated on. And the resource should be catching up as well.

And, definitely, I think regulators can be here to help provide oversight on ensuring companies are providing the right resource for families.

And, also, focus on that data collection part as well. You have to make sure ... Acknowledge children's data are collected, even though you don't really talk about it or say it, but definitely make that as the transparent part of the experience.

Genevieve Bonan:

Thank you so much. And we will get back to regulator's role in a minute. We have some time for some other questions.

Noura, if I could ask you, do you feel that it is difficult for the developers in your study to comply with applicable privacy laws? And if so, why?

Noura:

Yeah. I think our study showed that many developers are feeling the burden of compliance.

One of the main reasons is that many developers rely on third party SDKs they integrate in their own ads for advertising or analytics. And if you just open one of the websites for these SDKs, you can see that the privacy related information is scattered across many different pages across their privacy policies, terms of service, [inaudible 02:53:39] guides and everything.

And the amount of information that developers are required to process in these documents is a lot for them. And if you assume that many developers don't have the legal expertise or are not based in the US or the EU where some of these laws apply, then you can imagine how difficult it is for the average developers to actually comply with applicable privacy regulations.

In our studies, we found that many developers are actually independent developers, and so some of them also are part of small or medium sized organizations who are like ... They don't have dedicated teams for privacy compliance, and they don't have processes for privacy compliance.

Another major issue that they mentioned is that there is some kind of disconnect between developers and their legal teams. And, so, they assume that the responsibility of compliance is on their legal teams, while there are some tasks that developers should be responsible for taking care of.

For example, [inaudible 02:54:36] SDKs for compliance, many developers did not know that they had to do so. And when we discussed that with them, some of them said, "Our legal team did not tell us about that."

And, so, I think that there is a mix of process related inefficiencies within organizations. And, also, there is some legal background that many developers need in order to understand what their compliance obligations are.

And

Noura:

In our research, we found that many developers resort to some workarounds. For example, they copied some privacy policies from other popular apps, or they removed some of their apps from the Designed for Families program in the Google Play store, which is the program that has some specific requirements when it comes to child directed apps. And where their apps got rejected or excluded from this program, they decided that it might be easier to just say that our apps are not directed toward children. Yeah, I believe that there are many reasons why developers are feeling the burden of compliance and why we need to work on developing usable tools that help developers with their compliance efforts.

Genevieve Bonan:

We can't always blame the lawyers. But you mentioned in your presentation that only 31% of the developers that you studied have processes to screen the third-party SDKs that are running on their apps. How can the developers do a better job other than working with their legal team to vet these third-party SDKs?

Noura:

That's a good question. I think developers need some kind of guidance in order to understand what kind of information to look for in this documentation. For example, they can look for information on what kind of data SDKs collect or whether they share this data with other third parties. And I think that there are some requirements that are common among SDKs.

For example, recently if you go to SDK documentations, you would find that most of them have some kind of privacy methods or sections for privacy configurations. And they think that developers, who focus on developing child directed apps specifically, or apps that are subject to compliance with COPPA

or GDPR, should actually refer to these specific sections in the SDK documentations to see what SDKs actually expect from developers because different SDKs have different requirements and their ways of implementing these privacy compliance options are not always the same.

For example, some of them have only one server site configuration that a developer can just log in and just check a box and that's it. While others, have specific client-side functions that they have to configure, for example, collecting user consent, specifying that they're subject to compliance with CCPA or indicating that their apps are child directed. I think that developers can go and look for this specific information on SDK documentation, but I also believe that it's also a challenging task for developers. I think one of the main improvements that can be made in this space is to make sure that these configurations are consistent across different SDKs so that developers don't have to learn how each SDK is expecting them to configure basically their SDKs.

Genevieve Bonan:

Interesting. Thank you. I'm going to give you a rest and call up Kaiwen again. I was curious if you can talk a little more about the factors related to parents' perception and then their mitigation strategies of child safety in the smart home.

Kaiwen:

All right, thank you. Yeah, this part is where I don't get to talk in precision, but we did write it on the paper. There are fewer factors from both the parent side, the children's side, and the product side. From the parents' side, parenting style matters because some parents are really vigilant so they're more cautious about the safety risks versus some of the parents might be a little bit more relaxed. There's also the parent tech savviness and their trust in tech companies.

For example, if a parent are more tech savvy so they know ways, try to configure their devices and all that, resolve or and try to prevent some of the physical safety and digital risks. And for some of the brands that had news in terms of data leakage or being hacked, so parents would have less trust in those devices so they would avoid these.

And from the children's aspect, the biggest thing is their age and developmental differences. Parents definitely have more safety concerns with younger children and with those who have some developmental challenges. And lastly, I would say for the devices, like I mentioned earlier, that itself, if it's dangerous, let's say it produce heat or it has mechanic actions or elects certain CE or UL certifications, where some of them are cloud based or have no parental control whatsoever. These features are all red flags where parents watch out for. In all, those are the factors we realize that parents have been talking about, that affect their perceptions of safety and also mitigation strategies.

Genevieve Bonan:

Thank you. I'm glad you mentioned age because I was curious, are there age groups that you recommend starting with the smart devices? Is there any age that's too young?

Kaiwen:

That's a great question. Answer is, it depends on what is the particular type of device that parents intended to involve their kids with. Smart home technology, they can be used for entertainment, security protection, energy saving and all that. It's easy to think about something, let's say, with smart speakers, we see research have found how younger children from toddler to preschoolers try to interact with it, press the buttons or ask for songs or games and all that. I think that's a pretty basic first point of

contact between younger children and smart home for fun, entertainment, pleasure rather than utility or convenience.

That's usually for pre-teenagers or older children, I would say. Think about from a playful perspective, if we wanted to include younger children as part of this whole digital play or learnings and everything, we should think about age-appropriate design features and control to support parents, to help including their kids in these experiences. I think that's a good place to start and also considering how smart speakers usually serve as the central commander to control other smart home experiences via voice control and all that. Definitely smart speakers and younger children is a good place to start.

Genevieve Bonan:

Thank you. Noura, I could shift over to you. We have about five more minutes, and I have so many questions for both of you. But given that this research, that your research, Noura, compared developers' perspectives about their app's behavior with the actual behavior of their apps, did you find contradictions between what developers thought their apps did versus what they actually do?

Noura:

Yeah, we actually found contradictions between what they mentioned in their privacy policies and what we observed their apps do and we also found contradictions between their answers to our surveys and interviews and what their apps actually do. I think there is value in doing this comparative kind of research because it allows you to understand where misconceptions lie in developers' thinking.

For example, some developers thought that their privacy policies is closed, all their data or all their apps' data collection behaviors, but we found that their apps collected persistent identifiers with third-party SDKs. And this is, again, the reason why it's hard to deal with third-party SDKs for developers because sometimes they don't know what kind of information is collected by SDKs. And, also, we found some developers who believe that these kind of data collections are not their responsibility, it's the responsibility of SDKs who are actually collecting this data.

Yeah, we identified some contradictions and when we discussed that during the interviews, interestingly some developers said that they removed some of these SDKs from their apps in previous releases or they thought that they removed these SDKs, but it turned out that they removed the SDKs from some of their apps but not all of their apps. I think it's a combination of process related inefficiencies and lack of proper guidance for developers and also reliance on app store. Some of them said that we were not notified about these behaviors and so we assumed that our apps are fully compliant and there is no issue whatsoever with our apps.

Genevieve Bonan:

Of course. This question is for both of you, but, Noura, I'll start with you. Can you talk about what you would do to expand your analysis?

Noura:

Yeah, definitely. One future direction is examining other aspects of compliance processes that we have not investigated in this research. For example, future research could focus on examining how developers comply with the right to be forgotten or when users ask to not be contacted in the future and whether developers actually know where their users' data is stored in their systems and are actually prepared to delete the data when they're asked to do so.

Also, another future line of work that I would be interested in is, focusing on understanding how to empower users to verify claims made by app developers about their apps' privacy behaviors. Recently, app stores started adding data safety labels that allow developers to disclose what their apps data practices are, but the studies also have shown that this data is not verified, and users need help actually to verify that the claims made by developers are actually accurate and this would be another future direction to explore, how to empower users and, also, how to empower developers.

Also, I think it would be interesting to see how to explain to parents, in a language that they can understand, how their children's data will be used by third-party SDKs, specifically. And how persistent identifiers about their children might be used for behavior advertising or building profiles about their children. I think there's some kind information that might not be intuitive to parents when it comes to apps' behaviors or how third-party SDKs are handling data collected from their child users. Yeah, another interesting line of work.

Genevieve Bonan:

If you wouldn't mind, I'd love to hear Kaiwen; we only have a minute left. Thank you so much, Noura. Kaiwen, how would you expand your study?

Kaiwen:

Absolutely. I'll highlight two points out of the four we also write in the paper. First, is that our study focuses on parents' perspective, since parents tend to be the ones who adapt smart home and configure everything, in charge of children's access and use. However, parents' perspectives might not reflect on children's own perception and their smart home experience.

We definitely want to study how children conceptualize smart home safety and how both children and parents interact with both different devices, throughout their daily life, in their routines, and to observe if there's any tensions or conflicts or further issues, rather than only rely on parents' report. The second thing is that we really want to study families across different backgrounds like income levels, demographics, different tech savviness or educational backgrounds so we can get to see whether certain levels of challenges they experience, or certain groups have more or less challenges experienced on smart home technologies. Thank you so much.

Genevieve Bonan:

Thank you both. Now we will pass it off to the next panel entitled, "The devices are listening."

Tia Hutchinson:

Afternoon everyone and welcome to the fourth session of Privacy Con, entitled, "The devices are listening." My name is Tia Hutchinson, I'm a technologist within the Division of Privacy and Identity Protection, also commonly referred to as DPIP. I will be co-moderating this session along with my colleague, Bhavna.

Bhavna Changrani:

Good afternoon, my name is Bhavna Changrani and I'm an attorney in the Division of Privacy and Identity Protection. Thank you for joining us today.

Tia Hutchinson:

This session will examine the discrepancy between how consumers expect their audio data to be protected and how popular commercial products such as video conferencing applications, smart devices and other related products actually use the audio data that they collect. Our first speaker is Kassem Fawaz from the University of Wisconsin, Madison and he will present on the research and analysis that has been published in his paper, Are You Really Muted? A Privacy Analysis of Mute Buttons in Video Conferencing Apps.

Kassem Fawaz:

Well, hello everyone. It's really a nice pleasure to be here today. The next 10 minutes or so, I'll be talking about our analysis of the mute button in video conferencing apps and that analysis we performed last year in collaboration with our work with our colleagues at Loyola University in Chicago. Next.

As we all know, video conferencing apps like Zoom, Meet, Teams, Slack, Google Meet, and other apps, they do require access to your microphone and camera to function. Now access to a camera has been long studied and been recognized as a privacy problem. Now, most devices, like the one you mostly use, if you look at your camera, there's a green light that turns on or a red light that turns on whenever your camera is on. Now, if I turn off my camera, as I just did, that green light goes away. The reason being that green light is controlled by hardware switch. That hardware switch is controlled by an operating system switch. When you disable the camera in your app, that engages the OS control, which turns off the camera. Then we understand what's going on. That button controls what's going on at the hardware level and you have a reasonable expectation of privacy. Now, the second reasonable question or the second logical question is, what happens to the mute button? And that was the real motivation behind our study. Next.

Now to study the mute button in detail, we had to answer three sub-questions. The first, we needed to understand what are the user's expectations of the mute button? What do they think happens when you click the mute button and what do they think should happen when you click the mute button? Then we needed to compare that with the actual behavior of the mute button, which required runtime analysis of a bunch of binaries running on operating systems. And finally, we needed to study what are the ramifications of us accessing the microphone when the mute button is engaged? Next.

We conducted the study to understand the user's expectation of the mute button. We recruited 233 participants over the prolific platform to understand what do they think happens when the mute button is engaged? Do the VCAs still access the microphone? And whether the VCAs should still access the microphone when the mute button is engaged?

The answers to the second question were very clear: people think that VCAs should not access their microphone, when they're muted, which is what we all reasonably expect, right? When we click on that mute button, I can see a red cross or a red slash or whatever, which creates that there's some expectation that the access to the microphone is disabled. But when we look at the question or the answers to the first question, it was not very clear to them how VCAs do actually handle the mute button. Around half thought that the VCAs do not have access to the microphone when they're muted. Here we can see, in a snapshot, what do they think is happening, what they expect should happen and then what is actually happening. And this classical mismatch between users' understanding, user expectations and the actual behavior of a system or a device is where most of the privacy problems arise. And in our analysis, we took these studies and wanted to see what actually happens when the mute button is actually engaged. Next.

We studied a bunch of popular video conferencing apps, which you can see on the table over here, running on popular operating systems. We broke these applications into two categories. There are native applications that run directly on the operating system when you're running Zoom directly or Slack

or Microsoft Teams. And we looked at web-based implementation, stuff that required you to actually run them on browser like Google Meet, for example, you have to be on Chrome or another operating or another browser and then you run the video conferencing app from there.

We conducted a suite of planned time analysis to see, to intercept the communication between the app and the microphone APIs on all of these operating systems. And we identified three categories of behaviors. Next.

The first category of behavior is what we call the software bot and this and that was applied in Chrome based apps. When users on these web apps, like Google Meet, engage the mute button when they're on a call, that actually engages another control in the Chrome, which is the WebRTC mute button. In response, the Chrome browser will feed the web-based apps a vector of zero bots. In that case, the user is actually muted, the app has no access whatsoever to the mute button or to actual audio byte values. Next.

The next behavior, which is the most prevalent as we found in our study, is where the apps, and this is again when the user is muted, the apps will still keep access to the audio data, but they will not actively consume the audio bytes. This is when, for example, and this justifies for example, when you're muted, and you still see that indicator button showing that the microphone is being accessed. Because the app theoretically has access, they can access the mute button, they can access the microphone data, they can consume them, analyze them, but they choose not to do so.

And here, really, the user is at the mercy of these apps to well, behave. And all of this analysis was conducted on PCs, on Mac OS, on Linux and on Windows. When we performed that kind of analysis on mobile operating systems, which we didn't report in our paper, we found that the apps do actually consume the audio data and they do analyze that. And the theory was, apps needed to know the user is actually silent and if they're not silent, they wanted to alert them to unmute themselves. And in regular operating systems or in PC based operating systems, there's typically a flag that tells the app that the user is silent or not. In mobile operating systems, such a flag doesn't exist and that justifies why do they have to use the actual device. Next.

And the third behavior, which was very interesting, is that in some cases, multiple cases being Cisco WebEx, when the user is muted, the app continuously samples the audio data and they do process it. When we intercepted the API of microphone access, devices still flow to the WebEx app and the WebEx app was actually consuming them. And we found that behavior very interesting, so we dug deeper into that. Next.

And what was more interesting is, while the microphone data was being consumed continuously, periodic messages from the app and from the PC was leaving the device to WebEx telemetry services. We found a message leaving the device every one minute. We wanted to understand, is there a correlation between this data, being accessed to microphone data, and the network traffic leaving the device? And this data is obviously encrypted, so we looked at the encryption APIs on Windows and we tried to decrypt the network traffic and we were successful in doing that. And as you can see on the left-hand side of the slide over there, you can see some telemetry data of the audio that's leaving the device.

And you can see average statistic of the last minute of the audio data being developed, while the user is being muted and that is leaving the device. And then one might wonder, what is the big problem of this summary statistics of audio data leaving the device every minute? So, we did, what we can call, a thought exercise. We conducted a bunch of experiments where we had a bunch of activities running in the background and we looked at how does this data behave? Is there a correlation between this data over time and the kind of activity happening in the background? You can look at this graph over there, over the right-hand side, which I'm not sure is very clear, but you can see blobs of colors.

Each blob is an activity like cleaning, cooking, dog barking. And these blobs are very distinguishable and they're very separated in the space of this telemetry data, which is to say that the sequence of this telemetry information you can see over the right-hand side or the left-hand side for the seven minutes or 10 minutes, can fingerprint what activity you're doing in the background, which is very interesting. You might think that one minute average statistics might not tell much, it does. And that's consistent with previous research where average statistics of any kind of data over time to the information. Next.

We reported these findings to Cisco, and they investigated the issue and then two or three months later, after our disclosure, they released an official blog post of they acknowledged our findings, they confirmed them and they stopped continuously accessing data When the user is muted and they stopped sending telemetry information, which I think is a very good example of academic research having actual influence in the real world. Next.

There are lots of lessons to be learned from this. The most important lesson is, as with most of the digital systems and as you're going to be hearing later in the talks today, there is a mismatch between what users understand the devices are doing, their expectations and the actual behaviors of these systems and that system in understanding what the mute button of VCA.

VCAs should be clear about what the mute buttons should do and that should be indicated in privacy policies, because we found privacy policies are a bit ambiguous. And one of the more important lessons we learned, that there should be more native support to mute button. Similar to cameras, there should be hardware switches, but at the very least, software switches at the operating systems level to actually give the user assurance that mute button is actually engaged. And thank you.

Tia Hutchinson:

Thank you Kassem. I will now turn it to Bhavna to introduce our next speaker.

Bhavna Changrani:

Thank you, Tia. Next up we have Jide Edu from Kings College London, and he will be presenting his research paper titled, Measuring Alexa Skill Privacy Practices Across Three Years. Jide, you're up.

Jide Edu:

Hi. Hello everyone. My name is Jide, and I'll be taking you through our research on measuring Alexa Privacy Practices that we did across three years. Next.

Smart Personal Voice Assistant, which I'll be calling SPA, have become very popular systems, mostly due to the interactive technology. This has allowed users to easily interface with network appliances, as well as to consume all kind of online services using natural languages. Over a hundred million users now utilize SPA, like Alexa, Siri, Google Assistant, Cortana, Bixby and many more and they use this SPA to consume all kind of online services and also to manage the smartphone devices.

SPA also incorporate voice-driven applications which are usually developed by third parties, and they are referred to as skills in Amazon Alexa and Actions in Google Assistant. Next.

Like in mobile apps, skills play an essential role in understanding SPA capabilities by offering a wide range of services. And unlike mobile apps, skills do not run on user control devices, instead skills run in the cloud or in the server controls by the developer. The numbers of skills has multiplied in recent years. For instance, the Amazon Alexa skills ecosystem have grown from just 135 skills in early 2016 to over 100,000 skills by late 2020.

Along with the growth in skills, there is an increasing concern over the risks that third party skills pose to user privacy. Skills widen the attacks surface of this assistant as malicious actors might develop potential

harmful software that could affect the security and privacy of users. They request for permissions to access users' personal information and these permissions could have privacy implications. Recent stories have looked at issues in third party skills, including how developers have been publishing harmful skills, unjustifiably collecting data from the users and even eaves dropping conversations. Next.

However, it is unclear to what extent attacks permeate through the market. There are a huge number of skills which can make it difficult for SPA operators to vet this market. And this issue prompt us to ask these following open questions. How effective are SPA market operators in helping protect us? And one of the key features that need to be considered when answering this question is the strong dependency SPA holds with the cloud. Skills are hosted on remote web services controlled by the skill developers. This make it easy for developers to modify the skills functionality even after the skill has been published. And we focus on the following sub-questions, has the overall state of affairs regarding data practices in third party skills, ecosystems improve over time? Is the collection of customer information explained better nowadays and what influence changes over time? And has there be any improvement in the review and certification process of this third-party skill? Next.

To answer this question, we design a methodology to perform a data practices measurement which offer an independent assessment of the skilled marketplaces. Amazon operate different online marketplaces, which covers different 11 countries. That is the US, UK, France, Australia are among these marketplaces. What we did was to build a web scraper that scrape these online marketplaces and extract meta data from these skills. And in all, we collected three snapshots of market segments, one in May, 2019, one in July, 2020, and the last one in April, 2021. And we characterized every skill attribute we collected using their developer information, the privacy policy, the permission they collected, their API, the skills name descriptions and all of that attribute as well. Next.

There is need for one-to-one mapping between the data action collected by these skills and also what the developers did in the privacy policy that we're collecting from these users. We perform traceability analysis by looking at the privacy policy to understand what data these skills are collecting from the user and also the data that they disclosed to the users that they're collecting. Depending on how well traceable these data practices are disclosed, we classify it as either being broken, if there is no proper connections, or partial if they are not disclosed, but they are partially disclosed, and also complete if the information these skills are collecting from users are matched with what is disclosed in that privacy policy. Next.

For example, looking at how developers disclose their data practices in 2019 and in 2020, we found that developers disclosure practices in 2019 was bad and even 2020 was even worse. For instance, 40% of developers with skills that ask for permissions have all their skills showing complete traceability. This implies that all the skills statement in the privacy policy clearly state and justify their requested permissions. This is lower than what we saw in 2019, which is just 54%. And same thing for the skills with broken traceability, where we see 35% in 2019, compared to 51% that we see in 2020 and the key takeaway from this analysis was, developer disclosure practices was very bad in 2019 and even worse in 2020.

Based on our findings, at the head of 2020, we made the responsible disclosure to Amazon starting from mid-August, where we reported our findings, and we give them access to the skills that we found have proven traceability and Amazon confirmed receiving an email from the skills stores that they are taking actions.

In 2021, we measured the effect of the responsible disclosure we did in 2020. And where we see that 246 skills with broken traceability reported no longer posed a threat to users and overall, 256 skills have been taken down or have their permission removed by Amazon. But overall, we still see some of these reported skills are still posing threats to users and that's why the responsible disclosures, for instance,

we see 29% of the developers still having broken traceability, which the results shows that wider improvement in the traceability. There are still skills with privacy issues across the market. Likewise, it shows that while Amazon could benefit from more actionable mechanism like the one we did, they still have to properly implement some of our recommendations to ensure that the traceability could be better improved and likewise... Next.

Likewise, we also try to measure the effect of new skills on traceability to see whether the traceability issues we found are only associated with the old skills in the market. From our findings, we also see that out of the new skills added between 2019 to 2020, we saw 518 with complete traceability and there are still 114 with partial traceability and there are 364 with broken traceability. Similarly, to what we also found in 2020, even though the number has reduced a little bit, we still found 76 skills, many other skills with broken traceability and also 67 skills with partial traceability.

And one key takeaway is, though there is an improvement over the years, we still found many other skills having broken traceability, which shows that there are still room from improvement when it comes to how Amazon is doing the review process. Next.

To check our analysis for that, we performed something we call interrogative analysis in which we try to interact with some of these skills and see whether some of them are actually collecting personal information per interactions. And there have been existing research, name skill bed that have done something similar and we just leverage their automated interactive tools to try to interact with this skills case.

And from our findings, we discover 65 skills collecting personal information from users. And when we tried to check their traceability, we discovered 57% have broken traceability, 5% have partial traceability, and 30% have complete traceability. Out of the 57% with broken traceability, we discovered that 70% of them don't even have a privacy policy at all, to justify why they're collecting these permissions, which shows that Amazon is not actually giving priority to skills collecting personal information via conversations. I think we discovered that more priority has been given to skills that are asking users personal information via the Alexa API. Next.

Yeah, at the end of the longitudinal measurement study, we see that Amazon skill vetting process has improved over the years, though it's still not good enough. And we discovered that stricter scrutiny is given to skills that are collecting data via the Alexa API and none is given to those that are collecting personal information via conversations. And also, we see that even though skills could have a good traceability, they can still be over-privileged. And also, serval factors influences the traceability between the skills and the actual data practices they are collecting. And also, a responsible disclosure that we did, have a positive impact on the overall skill ecosystem. That's all. If you need more information, you can just check the paper online. Thank you.

Bhavna Changrani:

Thank you Jide. Next up we have Umar Iqbal from the University of Washington, and he will be presenting his research paper, Your Echoes are Heard: Tracking, Profiling and Ad Targeting in the Amazon smart speaker ecosystem.

Bhavna Changrani:

Umar, you need to unmute yourself.

Umar Iqbal:

I'm sorry. Yeah, so thank you so much for the introduction. My name is Umar, and I'm a postdoc at the University of Washington. Today, I'm going to talk about our research on primarily tracking, profiling, and ad targeting in the Amazon smart speaker ecosystem. Next slide, please.

So, before I dive into the privacy issues, I want to briefly explain how smart speakers work so that we are all on the same page. So, in a typical smart speaker interaction, the user gives a voice command to the device, like Amazon Echo, which records the raw audio and sends it to the voice assistant, like Amazon Alexa, in the cloud. The voice assistant then extracts the speech from the raw audio and processes it to understand the user command. If the command is directed to the voice assistant itself, it completes the task on its own. And in the case the command is directed to a third-party skill or an application, it is sent to the third-party skill server for completion. And after a task has been completed, the audio output is sent to the smart speaker, where it is played out to the user. Next slide, please.

So smart speakers provide a lot of convenience, and they're used by millions of people throughout the world. But at the same time, they process voice input, which can leak users' personal information. For example, tone in the raw audio can be processed to infer several physical and psychological traits of the speaker, like their age, their health, and their weight. Similarly, the processed raw audio in the form of transcripts can be traversed to extract personal information. For example, the transcript for command to a banking skill can leak users' financial information. And even the metadata generated by the voice command can leak sensitive information. For example, imagine a case where a user is interacting with, let's say a symptom checking skill, it can leak their medical condition. Smart speaker vendors like Amazon are aware of the presence of this information in user's voice, and they have filed patents to use that information for ad targeting.

What is even more concerning is that these are not just plans. Some of these patents have actually been operationalized in other Amazon products, like Amazon Halo, for instance. So now, considering the presence of sensitive information in user's voice, you would imagine that there would be more transparency in its collection and usage. But unfortunately, that is not the case, and I think it can be argued that even in a case where Amazon provides more information, users may not fully trust them to be transparent. Next slide, please.

So basically, to address this problem we built an auditing framework that brings transparency in data collection and usage in Amazon smart speaker ecosystem. The main goal of our framework is to provide independent and repeatable audits on unmodified off-the-shelf devices that eliminate the need to trust the platform. But achieving that goal is challenging, because smart speakers, unlike mobile phones and desktop devices, are closed box devices.

What I mean by this is that they do not provide any interfaces to measure precise data collection on end users. However, I think there's still an opportunity for us to infer broad data collection and usage factors. Basically, we can still infer if user data is collected and used for purposes that go beyond essential functionality offered by smart speakers. In our framework, we pick online advertising as a proxy of non-essential functionality because of two main reasons. First, data is most dominantly collected on the internet for targeted advertising, and particularly in case of Amazon, they have patents to use the smart speaker interaction data for ad targeting.

We measure data collection and users in online advertising by first leaking the data, and then looking for its usage in the ad content, and also through the change in advertisers bidding queue. Next slide, please.

So, the basic idea, as I just mentioned, is to leak data, and then look for its collection and usage. We leak data by installing and interacting with these skills on Amazon Echo. And as we're doing that, we intercept network traffic on a custom router to directly capture data collected by Amazon and other parties. And after that, we try to infer if the leak data is used for ad targeting.

We analyze ads in two places. First, the ads that are directly served audio ads, basically data directly served on the Amazon Echo, and also the web ads that will be served to a smart speaker user. And we capture those ads. We basically link the Amazon account that is associated with the smart speaker to a browser instance, and then start visiting popular websites, and collect the ad images and also data bits. And we repeat this process by installing multiple skills from nine different skill categories to establish statistical confidence in our inferences. And in addition, we also establish a baseline by repeating this process on an Amazon Echo, where we don't install or interact with any skills. Basically, don't return any data. Next slide, please.

We first analyze the data collection on Amazon Echo. The figure that you see on the left side of your screen shows the data collection by Amazon and third parties for different skill categories. We can see that among all of these categories, most of the data is collected by Amazon. And the reason behind this behavior is because Amazon mediates all transactions between skills and users. And even in a case where a skill does not want to share any data, they will have to share it because of the design, basically how all of these devices are designed. By looking at this figure, we can also make an observation that a significant jump of traffic, 8.3% to be precise, is directed to advertising and tracking services, which includes services like Megaphone and Podtrac that specialize in audio. Next slide.

So now, we analyze if Amazon influences user interest from their smart speaker interactions. To do that, we directly requested our data from Amazon for all smart speaker user profiles. And in the data returned by Amazon, we noticed that Amazon inferred user interest for several of our profiles. For example, for the smartphone profile, Amazon inferred that the user is interested in electronics, home, and kitchen. And in case of the fashion profile, they inferred that the user is interested in beauty, personal care, and fashion. Next slide.

So, after the user interests are inferred, we analyzed if they're used by Amazon to serve targeted ads. And in our analysis, we noticed several ads that were personalized and targeted to smart speaker profiles multiple times. For example, the dehumidifier ad that you see on the left side of the screen appeared seven times only on the health and fitness profile where we installed multiple air quality related skills.

And we saw similar trends for the audio ads. For example, we noticed several clothing ads from fashion brands in fashion and style profile, just like the one that see on the left bottom corner of your screen. And in addition to this ad content analysis, we've also analyzed advertiser's bidding behavior, because the ad content analysis is somewhere subjective. The figure that you see on the right side of your screen provides the advertisers bidding behavior for different smart speaker profiles. And you can see that the advertisers paid much higher for Amazon Echo user profiles as compared to the vanilla baseline where we don't leak any data. And in the majority of these cases, we found that the differences in bid values were statistically significant as compared to the baseline. Next slide.

So, after analyzing Amazon's data collection and usage, we checked if these practices are disclosed in Amazon's policies. For data collection, we found clear acknowledgement from Amazon that they collect smart speaker interaction data when users interact with Amazon. But for data usage, we did not find any explicit acknowledgement, or even denial that smart speaker interaction data is used for ad targeting across a range of disclosures from Amazon, which includes their privacy policy, Alexa Privacy Hub, and Alexa devices. The only disclosure we found was a public statement from an Amazon representative to the New York Times, which stated that Amazon does not use voice recordings for at targeting. This disclosure seemed to be inconsistent with Amazon's actual practices, because we clearly found that Amazon inferred user interest based on their interactions as per their own data. And we also saw users in ad targeting.

Recently, in a response to our research, Amazon acknowledged this, and basically, they acknowledge that they do use smart speaker interaction data for ad targeting. And I think one explanation for this inconsistency could be because Amazon might be trying to make a distinction between the direct use of raw voice recordings, as compared to the information derived from voice recordings, like the transcript or telemetry data. But that distinction might not be meaningful to the users. Next slide.

So next, we precisely evaluate that we try to understand user perception of Amazon's statement to the New York Times. And to do that, we conducted a user survey. And our study indicated that approximately 63% of the users do not expect Amazon to use their voice recordings, all the information derived from the voice recordings for ad targeting. And that was based on their interpretation to the New York Times statement. And we also found out people were uncomfortable with the use of voice recordings, or the information derived. Next slide.

So, in conclusion, we identified that Amazon and third parties, which includes advertising and tracking services, collect Amazon Echo interaction data. We also identified that Amazon infers user interests from their Echo interactions and uses those interests for targeted ads. Our findings also indicate that interest inference and their usage is potentially inconsistent with Amazon's disclosures. And lastly, our user study indicated that people do not expect Amazon to use their voice recordings or derived information for ad targeting, and they are mostly uncomfortable with that. At the end, I would invite you to visit the paper website to read other interesting details about our research, which I didn't get a chance to talk about here. And please feel free to reach out to me if you want to know more about our research. Thank you so much.

Bhavna Changrani:

Thank you, Umar. We'll use the rest of the time that we have to ask each of you some questions about your research. Umar, since you just ended, I'll start with some follow up questions. To your knowledge, can users opt out of allowing Amazon to collect voice data to infer interests and serve targeted ads?

Umar Iqbal:

So, in terms of user controls, Amazon provides a generic option where they can opt out of all targeted advertising from Amazon. Basically, as a user, you have to find that option. It's not apparent when you are, for example, setting up your smart speaker, or setting up your Amazon account, for example. And I think, to answer your question precisely, there are no specific controls that you can use to disable ad targeting based on your smart speaker interaction.

Bhavna Changrani:

Thank you. Do you think we're at a point where voice recognition analysis will continue to grow to include tone analysis to better understand the emotional state of a user?

Umar Iqbal:

Absolutely. I think we cite a large, I believe it's a book in our study, which basically highlights the information embedded in user's voice. It basically says that you can easily infer some basic characteristics like the gender, the age of the person, but it also indicates that you can measure their height, their weight, and other characteristics from the voice, along with a variety of health issues and psychological traits. So, I think it's possible at this point. I'm not a machine learning researcher, but I think it's... Basically, that's the caveat in my answer, but I believe it's certainly possible.

Bhavna Changrani:

Thank you. I'll turn it over to Tia to see if she has any questions for you or the other speakers.

Tia Hutchinson:

I don't have any further questions for Umar at this point, but I will move to Kassem. I have some questions about video conferencing apps. So, to anybody listening, they might be concerned that they're basically depending on the integrity of the company to not send their voice data to possibly third parties or advertisers. So, from your perspective, would you recommend that they perhaps use the web version of video conferencing apps in order to better protect their voice data, or maybe disabling the microphone themselves within the hardware, which may be dramatic, but maybe they're that concerned about their privacy. What do you recommend for consumers?

Kassem Fawaz:

Oh, that's a very good question that we hear a lot. I think users should be using whatever gives them the most assurance, and there's a spectrum of possible things they can do. At one end, they can use web-based apps, which have better guarantee in terms of the software mute functionality, but not all of the apps are usable in their web versions. I tried to use some of the apps in the web version, they are pretty unusable. The next step is using an operating system-based switch. I think Microsoft has this powered toy for a microphone where it can just disable microphone access completely to any app that's running. Recent versions of Android, you can look at the settings menu and you can disable the microphone option as well.

The third extreme is a hardware mute button. If your microphone has a harder mute that you click it and then nothing is flung through the wire to there. So, there's a spectrum, and people have their own valuations about usability, privacy, and usability and privacy tradeoffs, and everyone would use the solution they're more comfortable with, but they should definitely do something better than just relying on the mute button that exists. And that's my recommendation.

Tia Hutchinson:

Yeah, that makes sense. My next question, I was interested in the fact that Webex actually replied to your research, and they tested it, they confirmed what you saw, and they decided to zero out the telemetry data. Do you have any plans to test to confirm that this occurred?

Kassem Fawaz:

We do not. We're taking them for their word at this point, but we're trying to follow up in terms of behavior for apps on mobile operating systems, because our initial investigation shows that it's bigger of a problem there, because the audio bites do flow to the apps even when you're muted, which we didn't find in most of the PC based ones. So, more work is pending on the mobile side, and not just for audio, for other kinds of information that we might be analyzed locally.

Tia Hutchinson:

Okay. Bhavna, do you have any further questions?

Bhavna Changrani:

Not for Kassem, thank you. Okay.

Tia Hutchinson:

Okay. Go ahead.

Bhavna Changrani:

I have some follow up questions about your paper. Your paper mentioned that you reached out to a skill developer team at Amazon to share your findings, and I was interested to learn how that experience was. Were they receptive about your findings? Did they appear interested in updating their privacy policies to reflect your findings, or to address your findings?

Jide Edu:

Is that for me, right? Yeah. So, when we reach out to Amazon, and they said they wanted to set up a meeting with us, so that we gradually demonstrated what we have found, which they did, and we showed them what we have found, and we referred to them some of the data we have collected and the analysis we did. So they told us they are going to work on it, and after some months, we actually asked them, and they sent us an email that they have started working on it, though we don't really know to what extent have they actually implemented what we recommended, but they told us they're working on it, and we believe them, because from the analysis we did after a couple of months, we saw that some of these skills have been taken down, but we still see a lot of skills with privacy issues.

Bhavna Changrani:

And do you plan to follow up, either with additional testing or research sometime in near future?

Jide Edu:

Yeah, as part of the follow up, that was when we actually did, to try to understand the effect of the disclosure we did, and where we ought to see that some of the skills are still having privacy issues. Maybe we can still follow up later on to see how, or to understand how they implemented our recommendations, but we see that the traceability is getting better over time, which is the good news for us, yeah.

Bhavna Changrani:

Thank you. Tia, do you have any additional questions for our panel?

Tia Hutchinson:

I do. Jide, I have a couple more questions. From a consumer perspective, it's kind of difficult to expect that they're going to read hundreds of pages of disclosures to see what permissions have been requested, and what permissions have not been. Do you have any tips for the average consumer or what they can better do to make sure they're protected? Should they stay away from these apps that have poor traceability, or what would you recommend, based on your research?

Jide Edu:

Yeah, when we did our research and we discovered this finding, we thought it would be quite useful to have a tool that users can use, irrespective of their technical background, that they can just use that to understand the traceability and analysis of the skills they wanted to use. And then we actually developed a web application in conjunction with the Information Commissioner's Office of UK. So, we have these tools online, which we call SkillVet, that users can use to just analyze the traceability of their Amazon Alexa skills. It tells them whether it's broken, or partial, or complete, and it actually tells them

all the sentences or what statement have the privacy implications, or where is the data practices disclosed in the privacy policy. So, we have an open tool to support users to better understand what data practices these skills are collecting from them.

Tia Hutchinson:

That's great that there's a tool out there that users can use in order to confirm the level of traceability. My final question, I know you mentioned the conversation-based skills, which I think is interesting, because we always look at privacy disclosures, but then there's stuff that happens when we're just conversationally talking to Alexa. From your analysis, if a user decides, "Hey, you don't need to know my name in order to tell me the weather," for example, can they say no, or pass, or say something that basically denies them sharing that personal information and still receive whatever skill they're trying to receive?

Jide Edu:

Currently, as I said in the findings, the conversational base is not well scrutinized compared to the API base, and the conversational based ways of accessing, or data practices is sometimes not direct. In skills, you might initiate the conversation with the skills, and the skills might say, "How good are you," indirectly. So, you can decide to give the skills the wrong date of birth, but some consumers don't know that they just mentioned their age, and the most challenging thing is it's quite difficult to revoke such... If you grant such informations to those skills, there is no way you could go back and say, "Let me delete that informations," it's their conversations, it's not their API that you can withdraw your consent, so it's quite a bit challenging. And yeah, I would just say users need more awareness to know what informations to disclose, and what not to disclose. But yeah, the current way it's set up is quite difficult to withdraw your consent, especially when it is conversational based.

Tia Hutchinson:

Yeah, that makes sense. Bhavna, I think that's all the questions I have for the panel.

Bhavna Changrani:

Thank you. I'll offer a question to the panel as a whole, and someone hopefully take it. Is there any future or upcoming research that you can share with us? I'll pick on someone if I need to. Umar.

Jide Edu:

Yeah, let me go first. Yeah, from this specific analysis we did with the voice assistant, we also extend the same analysis to chatbots in messaging platforms, where we discover a lot of interesting findings as well. For instance, we discovered that in Discord, out of like 58% of those chatbots are asking for administrative permissions, administrative privileges, and only 4% of them actually have a privacy policy. And the most interesting thing is none of them have a complete traceability to justify why they're collecting these, and it's kind of a different ecosystems, and Discord were like, "Oh, it's let for users to actually find out whether they wanted to share their data with a third party chatbot or not," and it's not their own responsibility to do that for them. It's kind of like who has the responsibility to actually take care of these issues. So, is it the users, is it third party developers, or is it the platform provider themselves? Yeah.

Bhavna Changrani:

Generally, when a consumer's using that bot, and their understanding is that it's the website or the service that they're trying to utilize. Is that correct?

Jide Edu:

Yeah. What Discord is saying, "No, we are not the developer of this third party chatbot." So, it's now let for users to decide whether they want to use it or not, and whether they wanted to disclose their permissions or their data with this third party chatbot or not, that it's not our responsibility. We don't develop that; we don't have control over it. So, it's kind of like challenging, a little bit tricky for users.

Bhavna Changrani:

Great. Kassem or Umar, do you want to share if you're working on any upcoming research?

Kassem Fawaz:

Yeah. As I said, we are working on extending our work to the mobile operating systems, but not just sticking to the audio modality, to other modalities. The interesting thing is, in the old days, you used to think that privacy threats happen, or privacy incursions happen when data leaves the device. Now following up on your previous question to Umar about what kinds of information can be realized from the audio data, all of this analysis can happen on the device itself. They can identify emotions, they can find tone, they can find all of these things by just employing a large suite of machine learning models on the device itself. So, data doesn't have to leave that device for a privacy problem to happen.

Privacy problems can happen on the device, and then results of this analysis can leave the device, or even you can be [inaudible 03:56:22], your treatment can change according to this analysis, and that can happen on the device itself. So, ad personalization doesn't have to be a cloud process. That can happen, like a bucket of ads, and then do the personalization locally. Now that opens up an interesting question. Is this a privacy problem or not? If the data's not leaving the device, and you're getting the personalized treatment based on your data, but on the device itself, so what does that mean in terms of privacy? So that's the question we're really interested in, and we're investigating, we're digging into these apps and understand what's going on.

Bhavna Changrani:

That sounds incredibly interesting. I'd like to see where that leads.

Umar Iqbal:

Yeah, but I think it's really interesting discussion. So, I think as far as I'm concerned, I'm looking at a continuation of this project, and other things related to audio modality, and also health related data. But I think Kassem raised an interesting point about... I think he has two interesting points, one about local processing of data on the device, and harms initiated by it. I think it is important that we study both of these, especially the personalization that would result, and local processing, because I think the definition of privacy needs to be brought in a little. It needs to include issues like fairness, bias, and other things. So basically, if we include those issues, then I guess even the local device processing becomes important.

Bhavna Changrani:

Tia, do you have any other follow up?

Tia Hutchinson:

I think I'm all good over here.

Bhavna Changrani:

Great. I know we are scheduled to take an afternoon break. We're supposed to come back at 2:45 for panel five on augmented reality and virtual reality.

Erik Martin:

All right, great. Hello everyone. Very excited to do this session on augmented and virtual reality. We have some great speakers today to share their research. This is a topic that is a particular interest to the FTC. You heard the Chair mention it this morning, as well as our Chief Technology Officer Stephanie Nguyen. And given that this is sort of a technology that's still in the nascent stages, there is just a tremendous amount of unknown territory in terms of privacy implications, potential risks to consumers, and things that we at the FTC are also just trying to be smarter about. It's also interesting to track the trajectory of this technology. It started largely out of the gaming industry, gaming use cases, but now it's of course being used in education, healthcare, fitness, all kinds of other training applications. And so, the implications are far reaching, and of course, it includes noting more kinds of data being tracked with these kinds of devices. Eye-tracking, body movement profiles, muscle, and EMG signals, and sensing emotional state.

And then greater data persistence is another concern, as we're seeing more interoperability or connection between different kinds of virtual experiences and ecosystems, and just more kinds of virtual complex environments, which can be tricky for consumers to navigate and understand, and often is hard enough in sort of 2D environments to understand how and where your data is going, much less in some of these virtual spaces. And so, we're very excited to see more research in this space. And to start, we're going to hear from two of our speakers here, Jingjie Li, University of Wisconsin-Madison to talk about Kaleido, a real time privacy control for eye-tracking systems. And then Rahmadi Trimananda from the University of California, Irvine, to share research on auditing network traffic and privacy policies in Oculus VR in the Oculus ecosystem. So, I will hand it over to Jingjie to start us off.

Jingjie Li:

Hello. Hi everyone. I'm Jingjie Li, from the University of Wisconsin-Madison. Today, I'm very excited to present our work Kaleido, which is a privacy control [inaudible 04:01:18] for eye-tracking systems in real time. This work is done with researchers at UW-Madison and UCSD, and it was published at a USENIX security symposium last year. Next slide, please. First, why should we care about eye-tracking? Eye-tracking is an emerging interaction interface. Next, please, please. In eye-tracking systems, users' eye gaze are continuously checked by the cameras, regarding where they are looking at on the video scene. For example, the 3D world in VR. Such tracking enables hands-free interactions and applications, such as social avatar, foveated rendering, and event triggering. It is now more pervasively equipped in the coming commercial AR and VR platforms. Next slide, please.

But why should we worry about using eye-tracking? Before going deep into the privacy concerns, let's first see how does the eye-tracking data look like. Let's look at the figure, where I plot one user's eye gaze data as a heat map when they are browsing on an online store. We can see the two major clusters in the dense orange that represent users' attention. The eye gaze data can be [inaudible 04:02:35] into two categories, fixation and saccade. Fixations are eye gaze located close to each other, and it's associated with users' visual attention. Saccades are eye gaze traveling fast from one fixation to another. Users' fixation [inaudible 04:02:53] is associated with the regional interest in a visual scene, for example,

in this case a web component. [inaudible 04:03:00], they can analyze the spatial distribution of such eye gaze absolute positions to note about their attention, even subconsciously. Also, they can associate the aggregate statistic of eye gaze distribution with a lot of traits. Next click, please.

These traits include psychological traits and physiological traits, such as implicit interest, cultural backgrounds, personality traits, health status, like Alzheimer's or vision condition, and biometric identities and so on. Next click, please. So, this challenge motivates our design of Kaleido. There are some previous proposals to protect eye-tracking data by offline aggregation, feature [inaudible 04:03:52], and adding noise on the eye-tracking features before releasing data to untrusted applications. But if we do so, this will break the real time utility properties, and it will require aggregation of the raw data. But the users may not be so comfortable with this aggregation. Instead, we propose Kaleido, which sits as an intermediate privacy filter of eye gaze in real time in the local eye-tracking platforms. We accept raw streaming eye gaze data, and add noise on each data online. This privacy filter relays noise data across the trust boundary to the third-party applications. Since we still keep the format of eye-tracking data, the application can still use the same API to offer utility in real time. Kind of noisy for the formal privacy guarantee called local differential privacy. As we said, it is seamlessly integrated with eye-tracking ecosystems, and we also have a [inaudible 04:04:51]. Further, we automate privacy configurations to save users effort. Next, please.

Jingjie Li:

Next, how can we define the privacy for Kaleido then using these formal privacy guarantees? First, think about that eye tracking is actually a location data, and the spatial information of eye tracking data actually is the primary source of a lot of sensitive attributes that enables attack and threats like [inaudible 04:05:25]. Plus, without the differential privacy framework called geo-indistinguishability, this framework was originally designed for geo locations privacy. Its noising will ensure that the results are indistinguishable for all pairs of input within distance r. On the other hand, giving privacy for just one individual eye gaze make no sense. As in the realistic scenario, it's a streaming data and we care about it, but it's also impossible and not realistic to offer privacy for the infinite stream without a fixed glance. So, we consider more realistic scenario, and we give privacy on any slightly window of a duration W within the infinite stream. And this makes sense, since in the AR and VR applications, the things are usually changing very fast. Next, please.

In this slide, we intuitively visualize Kaleido's effect using the eye gaze heat maps on the webpage. First, you can see the difference between different settings. On the left most scenario, Kaleido has no main being enabled in this scenario. You can see that the users is focusing on mostly on the bottom right corner of a web component. If we started to add privacy and Kaleido's privacy is characterized by price budget apps, we can see that the tension is jetted to the middle of the page. And if we start to add more privacy in the right figure, we can see that the user's intention is more scattered, and tech will be more confused. Next, please.

We also implement Kaleido in Unity. Unity is a mainstream AR and VR development platform. To evaluate how our system works in the world, we evaluate Kaleido in Unity actually again where users can use their eye gazes to direct laser and should zombie rabbits. Here, I want to show the two video clips with without privacy enabled, and with privacy enabled. Please help play the video with low privacy, and the other one on the left side is no privacy. Thank you. So, from the videos, we can barely see that there's significant difference on the utility impact. And in our user study, we test our game using a PC webcam eye checking set up during COVID period, and we evaluate five settings in anonymized and randomize order except the control knob setup where we allow users to adjust the privacy level during the game play. Next slide please.

In this slide, we report results in the subjective and objective metrics. We ask users to rate their enjoyment levels and we record their game score. In other words, the number of rabbits they take down during the game using eye tracking. We see that even with no privacy, the experience degradation is negligible, and if we even increase the privacy level, it posed minor impact. Next, please.

Also, we care about Kaleido's effectiveness against different attacks. Here show two examples of major attacks and the results of Kaleido. The first major attack is interest based where the attacker is trying to identify the outlier users with distinct viewing patterns per image scene by using clustering. This image scene may include the viewing natural image or webpage or some sort of human face. Such viewing patterns is associated with users' mental status, culture, or social backgrounds, and so on. On the other hand, we value the biometric attacks where the attackers trying to identify users' traits from biometric features during video sessions. These biometric traits include their identity, also, the vision tracking condition. Both of these two attacks will enable consumer surveillance. For example, delivering targeted advertisements using the user's implicit interest, for example, certain culture background. And the results show that attacker success can be brought to random guess at high privacy, and with low privacy, the tech can still be thwart greatly. Next, please.

Here, we summarize some our findings. Kaleido, as we said, it is a first system to protect privacy for eye tracking in real time. Moreover, we demonstrate how to deploy differential privacy by leveraging the semantics of eye gazes, which is an interactive application. Also, Kaleido can be seamlessly integrated into the existing eye tracking ecosystem and keep the original API, and we demonstrated in our Unity implementation as well. Kaleido is user facing, which provides a flexible control knob for users to tune their usability and privacy online. Next please.

Here, I want to talk about our ongoing work and roadmap to carry on Kaleido. We first want to understand the utility tradeoffs in varying eye tracking applications, and these applications may have very different semantics and requirements. In our first study, we developed a [inaudible 04:11:16] application, but there are other important applications such as foveated rendering. Foveated rendering is one application in AR and VR platform where user's attention area will be rendered with high resolution content rather than the other areas to save computation. Well, more than just utility and perceptions, this brings along another dimensions of trade off of computational cost. Another example is the social interaction where the eye tracking data is consumed by multiple users at the same time.

Then I want to study users' understanding of differential privacy for eye tracking data, specifically how users' privacy perception align with the theoretical guarantee and how users balance their privacy between usability in their adoption and use across different applications. Putting all of this together, I want to design effective privacy communication and control interface that helps user make better privacy decisions. Thus, the privacy of eye tracking and AR and VR is a rising area, and there are a lot of unknowns and underlying threats and solutions we can explore. I'm very happy to connect and collaborate with everyone and researchers who are interested in this area and here is my contact information. I'm also on the academic job market this year. Feel free to reach out and chat. Thank you. That concludes my talk.

Kassem Fawaz:

Awesome, thank you Jingjie. And I will turn it over to Rahmadi and possibly Athina since we have her on the line as well and are very glad that we do.

Jide Edu:

Thank you so much Erik. Hi everyone, my name is Rahmadi Trimananda. I'm a project scientist at UC Irvine in the ProperData Center. Today I'm presenting our work OVRseen, which is a system and

framework for auditing network traffic and privacy policies in Oculus VR. This is joint work with Hieu Le, Hao Cui, Janice Ho, Anastasia Shuba, and Athina Markopoulou. Next slide please.

Virtual reality is a newly emerging technology that has great potential and wide range of applications from games to education and healthcare. And Meta is currently a big proponent of this technology through Oculus and in particular, the Quest 2 headset is one of the most popular VR headsets. When we are using the VR headsets such as a Quest 2 headset, we are using a VR app and we're interacting with the VR environment in the app. For instance, when we're playing the Beat Saber game as shown in the slide, we are tasked to slash red and blue boxes that are appearing in front of us using two light sabers in our hands as we are following the beat of the music. Next slide please.

In the real world, however, the VR sensors on the device may collect body and motion data on top of our data that are being collected as well, such as personally identifiable information or PII. And on device ads are also coming that can be immersive and use sensor and biometrics info. In short, VR is the next big thing, but as far as the implications are not well understood yet. Next slide please.

This motivates our work in which we perform auditing on the system and policy and VR. As we can see on top, on the system side, multiple apps can run on a platform and together they can send data in the network traffic to various destinations. Down below, we can see that the privacy policies of apps and platform, they declare the data collection practices as mandated by privacy laws such as the GDPR and CCPA. Here auditing means comparing and checking the consistency between the actual data collection practices that can be found in network traffic as systems output with the declared practices that can be found in the privacy policies.

To enable this, we adopt an existing framework called contextual integrity, which is meant to assess information flow according to its transmission principle. Information flow consists of sender, data type, recipient, and subject. In this context, sender can be the app or platform. Data type can be PII. Recipient can be first or third party destination, and subject is typical of the user as a data owner. Transmission principle is basically the context of the flow, which can be the purpose of data collection in other contexts. Together, these elements form a couple and in this work we focus on the four elements highlighted in red, namely the sender, data type, recipient, and purpose. Next slide please.

So, suppose we wanted to perform this auditing for the Beat Saber game, and suppose we could extract from the network traffic the top pole that's shown on top. We call this on a network traffic site a data flow. Here the sender is Beat Saber, data type is user ID, the recipient is beat games as a first party destination. And so, we infer that this flow, the purpose is most likely for functionality. And suppose we could also analyze a corresponding collection statement from the Beat Saber's privacy policy as shown down below. And in this very example, the top elements extracted from both sides, they exactly match and so they're consistent. So, to summarize, auditing means extracting the data flows from network traffic, analyzing the collection statements from the privacy policies, and checking the consistency between the two. Next slide please.

Now to enable this in VR, we created and used OVRseen. The first part of OVRseen is network traffic analysis. Starting from the left-hand side, we can see that we start from collecting 150 most popular apps from both app stores. Oculus, the official app store, and SideQuest, the most popular third-party app store. We run each app for approximately seven minutes on the Quest 2 headset while we're collecting the network traffic using EnMonitor. Meanwhile, we also use Frida to bypass the certificate validation for the traffic and encryption. And this allows EnMonitor to collect and decry the traffic at the same time. Upon collecting the network traffic in a form of raw data, a pick up NG format, we convert this file into JSON format, and we extract the data flows. Each data flow consists of app or sender data type and destination or recipient, and then we performed two types of network traffic analysis, namely data types analysis to understand what data types are being sent in a network traffic and ATS or asset

tracking services analysis to understand what destinations are being contacted in a network traffic. Next slide please.

Here our results show that, on the left-hand side we can see a table that shows the 21 data types that are sent in the network traffic. They include PII or data types that can identify a user, fingerprint or data types that can be used to track a user mostly used by browsers, and most uniquely also found VR specific data types from the VR sensors. On top, we can see that the top destination entities include Oculus and Facebook as platform parties, and Unity as the most popular third-party library used by VR app developers. So, the takeaways here are that Oculus VR is a young ATS ecosystem whose data collection practices include sending 21 data types to various social or analytics tracking domains. And so far during our experiments, we did not find on device ads yet. Next slide please.

Now upon collecting the network traffic, we want to also perform the full auditing by comparing the data flows and network traffic and the corresponding collection statements in the privacy policy. And so, the second part of what we're seeing as shown down below in this diagram, it's called privacy policy analysis. We start by collecting privacy policies from all the platform, the apps, and also third parties. And early on, we found that a lot of apps actually did not provide a privacy policy, namely 27% of them including those in the official Oculus app store. When they have a pricing policy, we used our privacy policy analyzer to check the consistency between the data flows in the network traffic with the collection statements from the pricing policy. Our analyzer leverages two state art of tools, namely PoliCheck as shown in the red box, and policies as shown in the green box.

PoliCheck relies on the ontologies to determine whether the two sides are consistent or not, namely the data flows and the collection statements. These are data ontology and [inaudible 04:20:31] ontology. The data ontology for example, it maps the more specific terms that can be found and the data flows such as user ID, and the more generic terms that can be found in the corresponding collection statement, for example, device information. And so, it decides that the two sites are consistent if say the data flow contains user ID and the collection statement contains user ID or it can also consider that the two sites are consistent if the data flow contains user ID as a more specific data type and the collection statement contains device information, which is a more generic data type. Next slide please.

So, our results shown here, it shows that 70% of the data flows that we collected were inconsistent with the collection statements and the privacy policies, mainly for fingerprint and VR sensory data types. Upon investigating further, we found that a lot of privacy policies and apps did not reference platform and third-party policies, whereas a lot of the data flows were sent to platform and third-party destinations. And so, to confirm and to clarify, we conducted a second experiment by including these platform and third-party policies into the analysis by default and we found that 74% of the data flows became consistent. And so, the takeaways here are that many of the developers' privacy policies are missing. If they do exist, they're poorly written and they neglect referencing third party policies. Next slide please.

So finally, OVRseen also integrates this purpose extraction by integrating policies another state of year two. And so, we can extract the purpose and complete a couple from just three elements to four elements by translating the sentences extracted or and analyzed by PoliCheck into tech segments that are labeled with purposes by policies. We found that close to 70% of the data flows were for non-core functionality and only around 30% of them were for core functionality. And so, the takeaway here is most data flows were for non-core functionality and purposes such as advertising and analytics.

We also responsibly disclosed our findings to developers and Meta. We received mostly positive feedback from 24 developers saying that they're grateful for our findings and they're in need of any means or tools that can help them be more compliant with privacy laws in terms of their data collection practices. Unfortunately, we did not receive any response from Meta. Next slide please.

Thank you so much for your attention. The code and datasets are available. This work was also published at USENIX Security 2022 and I'm happy to discuss further and take your questions. Thank you.

Kassem Fawaz:

Thank you Rahmadi, that was a great presentation. All right, so to start here, I have a question for you Rahmadi, which is so you have this really incredible finding in your study where you show about 70% of the data flows from VR apps were not disclosed or consistent with the privacy policies, only 30% were consistent. 38 apps did not have privacy policies, fewer than 10 out of 102 apps that provided privacy policy explicitly asked users to read it and give consent when first opening the app. Did you see any patterns in terms of the kinds of apps that were the worst offenders or just in terms of that distribution of who was consistent versus not and who was providing notice and consent versus who was not?

Jide Edu:

Thanks for the question, Erik. I think in general; I think the developers and the apps, they don't really have bad intentions. Most of the time, they just neglect declaring things in their privacy policies. Mostly because a lot of the VR developers, they're mostly a really small firm, small company, right? So, they'll probably only have one or two people, sometimes more, but not like a big company with legal people who understand privacy policies. And so, the fact that they're not declaring things usually because they're aware that they're not trying to collect any data, but they're not aware that they're using third party libraries that are collecting data such as Unity. Unity is always used, not always, but heavily used by a lot of apps, maybe 70 to 80% of the VR apps.

Kassem Fawaz:

Gotcha. And I mean, there's sort of a steep fall off in the distribution of traffic going to third parties and Unity is, by far gets the largest share of that third party traffic.

Jide Edu:

Right.

Kassem Fawaz:

And do you imagine that might change in the future or what could change that dynamic? There's just not that many sort of other third parties that are used that widely by most of the VR apps that are out there? Or what's sort of the dynamic there?

Jide Edu:

Right, so two things that I need to mention as a context of our findings. So first of all, we only collected the traffic for seven minutes per app, so that's really a small amount of time. When you're playing with a VR app, you will lose track of time most likely and you would play for hours and hours so I would only imagine that you would probably find more third parties in that network traffic if you're playing for say two hours per app. And number two, this study was done back in April 2021 and at that time, the Quest 2 had said was just the state of the art and it was just released a few months before, so the ecosystem was just in its infancy. So, for now, there's a gap of more than a year. So perhaps today if one would like to collect network traffic, they would find more of the third parties in the network traffic.

Kassem Fawaz:

Gotcha. Is there a similar sort of situation, you're talking about the fact that a lot of these VR apps are built by very small studios and teams and you show the majority of the data flows are for non-core functionality, and is that just a function of when people are building these apps, they're not being particularly careful about all of the other, they're collecting more than they necessarily need to, or they're relying on libraries for purposes that they don't necessarily need to just because it's easier for a small team to do that? Or what's going on there do you think?

Jide Edu:

Right, so I think the reason why a lot of developers use Unity is because they provide this analytics service, right? Another game engine that people also use is Unreal, but as far as I know, I don't know today, but when we did our experiment, they did not have that service to provide that analytics measurements and stuff in terms of their users. So, I think they're relying, for example, on Unity to provide them with the measurements on the user side, how they perceive the game, how they use the game, how long, what they can see, the behaviors of the user and stuff like that to improve the game itself.

But again yeah, a lot of times, well, they did not realize that they should have declared the third-party libraries as well. Like hey, please read the privacy policy of Unity because we're using their service. There's also some others like PlayFab API, Log Me, those are also third-party vendors that provide analytic services. And so yeah, one main gap, one key gap that we found is that I think developers, they just didn't realize that they should have pointed users to those other privacy policies, or at least mention them in their privacy policies. And funny enough, some of them just write, this is our privacy policy, just one statement. We don't collect any data, but they're using a ton of third party [inaudible 04:29:21] libraries.

Kassem Fawaz:

Gotcha. And the analytics services for gameplay and monitoring.

Jide Edu:

Right.

Kassem Fawaz:

Athina, it looks like you have your hand up if you wanted to chime in as well. Oh, you're muted.

Athina Markopoulou:

Yes. So, I would like to add a comment about the projection for the future. So right now, the VR ecosystem is still young, and people not figure out how to do ad scale, but as ads get added, that may add another set of destinations where the data are flowing to. So far, the indication is that this thing is a first party tracking system, but that may change in the future. Thank you.

Kassem Fawaz:

Right. Well, and let me turn to Jingjie here. Well, actually this is sort of a question that can go to both of you, and actually Rahmadi in your paper and Athina, there was a study that was noted, revealed there's the feasibility of identifying users with a simple machine learning model using less than five minutes of body motion tracking data from a VR device. And Jingjie, I was curious, the plugin that you built and the technique that you used seems to have been really effective in terms of mitigating, preventing

identifying the user even at fairly low settings that you'd constructed. But if that technique could be used for other kinds of biometric data or you've done eye sensing, but obviously VR, AR, they're going to be collecting tremendous amount of other kinds of data and where you see that potentially being relevant.

Jingjie Li:

Yeah, thank you Erik. That's actually a very good question. So, the good thing about our work is to utilize the framework for differential privacy that have this, probably a guarantee that generalized across different texts. But we need to keep in mind that if we want to migrate this framework to other data, for example EMG, it's your muscle signal, and there are a lot of rising AR and VR devices, they try to use this EMG to allow user to action controls or to learn their motion intention. And these signals, they also convey a lot of information like the health status or say stress level.

But the issue is the research challenges, if we want to migrate the differential privacy framework used for eye tracking data to other biometrics, we could think about what's the proper semantics of this data. Since in our work, we associate our privacy, the privacy semantics and press framework with the locations, individual thing and it's well connected with how the eye tracking data is formed naturally. But in EMG, it may not necessarily be the same thing, so it's still an open question. How do we protect different data types? Do we need to customize for just protection level based on the different guarantees, and moreover, the different utilities, and we can do post processing to preserve and calibrate this utility in different applications. Thank you.

Kassem Fawaz:

Super interesting. And Rahmadi, in the research that you did too, could you speak more about the VR specific data you'd identified and for what purposes you saw being used and how you think that that might evolve as well?

Jide Edu:

Sure, thanks Erik. So, for example, they're collecting the VR play area. So initially, usually when we start using the headset, we are required to draw the area that we're going to be playing with in, I mean inside our house. So, the privacy implication of this is, for example, one can infer the size of the room or even the house that you're in and who knows, in the future this kind of data type can be used to infer your location, for example, or just to infer something about your home. And another interesting group of data types that we found was VR movement, which is a collection of a lot of sensors. Even last year, they were already sending gyroscope data, magnetometer, accelerometer, proximity sensor. All in all, they're trying to measure everything about our movement.

On one side, this is necessary because they have to ascend those data to keep the functionality of the game. How can we play the game or the app if the sensors are not detecting anything? But on the other hand, who knows what they can do with the data they're collecting? So, in the future, as you mentioned earlier, we cited an earlier research work on how one can predict or how one can identify a user by using the body motion data. From the way we walk, the way we see things, the way we gaze in relation to eye tracking for example, the way we move in a certain way, the timing and everything, I think all of them combined can identify a user.

Kassem Fawaz:

Interesting. And so far, you're not seeing much use for advertising purposes, but obviously that could change.

Jide Edu:

No, in terms of advertising, I think you brought up a really interesting point of view. Indeed, during our experiment, as Athina mentioned earlier, it looks like Meta is controlling mostly almost everything. Well, especially during our experiment, right. And then we heard the announcement from Meta. So, three things. The first one, we heard the announcement from Meta about their testing VR ads probably around mid-2021, around June or July, they announced that they were testing this ads framework that can show an ad on a virtual wall when you're playing with a VR game. But they discontinued that effort so there was no further announcement.

Number two, we also have seen a big push from Unity that developers can integrate ads or maybe a framework for ads to be shown in their game. So interestingly, one can integrate as a developer, the ads into the flow of their game. So, for example, as a VR user, we can play with the game and at some point, we can enter a room that's full of ads, and we should solve a mystery with respect to the ad and it could be an enjoyable experience. I haven't tried it myself, but I would imagine that this would be more appealing than just some ads that are showing during our TV show, for example, on a streaming service like Netflix for example.

Number three, so this is something that I've tried myself. There's a big push from the Meta side as well for the metaverse apps. And some of us, I think we know that there's this Horizon World, an app that's pushed by Meta to be used by hopefully a lot of users and in which you can create your own world, like a metaverse of your own, a universe. And then probably by now, there should be thousands if not millions of worlds there. And each world can be unique to yourself, we can invite other users to come to your world. Now, we have seen that these worlds can be used by companies to actually advertise their products. So, I visited for instance, the Wendyverse for the Wendy's burger. You can go to the virtual store, interact with people. I don't think you can eat any burger there, but you can interact with some machines there and then you can get a virtual voucher that you can exchange with real, just chicken nuggets unfortunately. But I think it's a good advertising too [inaudible 04:38:03].

Kassem Fawaz:

Yeah, and we're seeing something similar even with Roblox, which isn't VR, AR, but having sponsored experiences. Last question here for you Jingjie, what was interesting also in your study was you looked at player attitudes around when they could toggle up the privacy enhancement. And it seemed to be the case that neither enjoyment of the game nor the actual outcome, the scores of the games were badly impacted. And I was curious if you plan to do sort of more research around consumers' willingness or interest in using those features or ideally it would probably be the developers who just make this the default and what you think about that. Quick answer.

Jingjie Li:

Yeah, thanks Erik, actually what we are going to do now, because we have a lot of interest finding in our first study that though the users, they are willing to enable such control, but they still control it differently during the game play. Some of them, they plan to do calibrations and stay on certain pricing level in the beginning of the game. Some of them, they try to play around with it, that's make them feel more comfortable throughout the game play. But there's some still challenges to the users in understanding what does this privacy level mean to them, although it provides some rough guidelines. For example, how the high-end low privacy match to the interest, the users, they still demand more information. And there's also some recent study in understanding user's perception and expectation for differential privacy for say health record or smartphone data. And they see a gap, they see sometimes user, they're not really aware of it and their privacy guarantee. And this angle is very interesting in

Jingjie Li:

Our scenario is that our utility is evaluated in the perception and how to use a balance is between this unknown threat and how do we communicate with them, that is the work we are going to do now. How we offer interfaces allow them to better control, make better decisions. How do we let them understand these implications and do the trust between the platform and the users?

Erik Martin:

Great. All right, well we will stay tuned for that and a reminder Jingjie is on the academic job market, and I will now have the pleasure of handing it off to the next panel on interfaces and dark patterns. Thank you everyone for tuning in.

Gorana Neskovic:

Thank you, Eric. Welcome to panel six on interfaces and dark patterns. My name is Gorana Neskovic, I'm an attorney in the Division of Privacy and Identity Protection.

Min Hee Kim:

And my name is Min Hee Kim, an investigator from the Office of Technology Research and Investigations.

Gorana Neskovic:

Coined in 2010 by user design specialist Harry Brignull, the term dark patterns has been used to describe design practices that trick or manipulate users into making choices that they would not otherwise have made and that may cause harm.

If you would like to learn more about FTCs prior work on dark patterns, please visit our website where you can find a transcript of our Dark Patterns Workshop held in April 2021, as well as FTC staff report on dark patterns published in September 2022. Today's panel will explore how dark patterns impact cookie consent interfaces and how dark patterns may impact consumer experience across websites and mobile apps.

In just a moment, you'll hear from Hana Habib of Carnegie Mellon University and Johanna Gunawan from Northeastern University. Detailed bios of all of our panelists and links to their research papers are available on our PrivacyCon 2022 event page at ftc.gov. Without further ado, I will turn it over to Hana to start us off.

Hana Habib:

Thank you. Good afternoon, everyone. Today I'm presenting work on Cookie Consent Interfaces on behalf of my colleagues at Carnegie Mellon. This work was published at Chi earlier this year, so I encourage you to check out our paper. Next slide please.

Interfaces like this have probably become very familiar to you. Common reaction is to dismiss the interface as quickly as possible, suggesting that they suffer from usability problems. So, these seemingly

ubiquitous interfaces are used to meet regulatory notice and choice requirements such as those laid out by the privacy directive, GDPR and CCPA. However, prior work has found that patterns that steer users to less privacy protective options are prevalent in these interfaces, which kind of violate the sphere of these regulations. It's also important to note that these interfaces are commonly implemented through one of five consent management platforms or CMPs, which has somewhat standardized how they look. Next slide please.

So, while we know that the continued use of consent interfaces is not the ideal usable solution for capturing user consent, until other mechanisms kind of gain traction, we can at least explore how to make them better. So, we conducted a two-phase usability evaluation of consent interfaces. First, we inspected around 200 cookie consent interfaces from five CMPs, which identified key design parameters or ways that these interfaces could differ from one another. We explored the usability impact of these design parameters through a usability study in which we tested 12 design variants with users to evaluate their impact on six usability aspects. Next slide please.

While I'll skip over the details of our inspection of existing consent interfaces, which are described in our paper, this evaluation helped us identify seven design parameters which we observe kind of change across different implementations. So, we wanted to explore the impact of these design choices in our user study. And these parameters included things like how prominent the consent interface was, the path that led to cookie options that corresponded to different cookie categories. So, whether the choices were framed in a way that would make people feel afraid that they would be losing something, which you call loss subversion. The readability of the consent interface, the text within button options, the format of the cookie options that's linked from the consent interface and the process for changing a consent decision. Next slide please.

For our user study, we designed in between subjects experiment in which participants were exposed to a design variant. So, our baseline variant, which we call "best-practices," incorporated what we hypothesize would be the most privacy-protective, or usable option for each parameter explored. Specifically, it was a fully blocking design which made other parts of the website inaccessible. It included options for cookie categories in the initial screen through check boxes for different cookie categories. The notice texts was formatted in bullets to help with readability and the button text here detailed the actions performed by the buttons and it noted that users could change their decision through a persistent cookie preferences button. Next slide please.

Our best-practices variant also had this link show details which led to a version of the cookie preferences interface for all options for categories were laid out on the same page. Next slide please. For 10 of the study conditions, we changed just one aspect of the design from our baseline to be something that was less usable or less privacy protective so we could directly measure the impact of a particular design choice. We also tested a "Worst-practices" variant to explore the combination of the least privacy protective or usable design options that we explored. It was implemented as a banner which didn't block users from other parts of the website. It included loss aversion text that warned users that their experience may be affected if they didn't accept optional cookies. The text of this notice was formatted in a paragraph and the button didn't specify an action, it just said "Okay." Decision reversal was not mentioned at all in this banner. Importantly, options to cookie categories were sort of hidden within a link that was embedded in the text. Next slide please.

And this link led to a multilayer cookie preferences screen that had options for cookie categories and different tabs rather than a single page. Next slide please.

We tested our design variance with over 1300 prolific participants and the participants were assigned to a shopping task on a prototype website where they were exposed to one of 12 of our design variants. They were asked to then answer questions based on what they remembered from their experience, and

they were asked to look at the consent interface again and answer more survey questions. The median completion time for a study was about 16 minutes and participants are compensated $5. We analyze interactions and survey responses from about 1100 participants. Next slide please.

When looking at consent decisions participants reported making, we found that some design variables had more of an impact, while others, not so much. So, each bar in this graph represents one of our design variants and the colors here are different consent decisions. We found that changes to the paragraph or button text didn't really have much impact. However, Next slide please.

Participants were more likely to consent to all cookies and conditions that didn't include in-line options on the initial consent screen, such as the worst-practices banner I showed a few slides back. Next slide please.

We also found that the terms used to describe standard cookie categories, which we drew from the ICC UK Cookie Guide cause confusion. Less than half of participants selected the correct definitions for performance cookies, which are cookies that help measure and improved website features. And only 16% selected the correct definition for functional cookies, which are cookies that help personalize the website services. Next slide please.

In one of our design variants, the cookie consent interface was simply just a cookie preferences button in the bottom right side of the screen, which was modeled off of an option that's provided by the CMP OneTrust. We found that it was largely ignored when it wasn't accompanied by a banner or a pop-up interface. However, when it was paired with such an interface, it enabled participants to change their consent decision. We found that 82% of participants who were in that best practices condition said that they would use that button to change their consent decision. While only 45% of those who were instructed to visit the cookie policy and didn't have that button available to them, said that they would go to the cookie policy to do so. And we didn't find a significant impact by removing that instruction text within the consent interface suggesting that the button on its own is effective. Next slide please.

As I mentioned at the beginning of this presentation, consent interfaces pose a burden beyond their poor usability. There's a considerable cost to reading consent interfaces, comprehending available options, and making a decision. The browser-based consent management mechanisms, including browser extensions, have the potential to alleviate some of those burden. However, until these consent mechanisms become widely adopted, it remains important to improve the usability of existing schemes, particularly when it comes to dark patterns. Next slide, please.

Thank you. If you have any questions, please feel free to reach out and email me.

Min Hee Kim:

Great, thanks Hana for that peak into users and their perceptions about cookie interfaces. I think that's particularly interesting given the volume of these notices that we encounter, some of which as you mentioned, don't really give us much of a choice. And so next we'll turn to Johanna to tell us about dark patterns across web and mobile environments. Johanna.

Johanna Gunawan:

Thanks Min Hee. Yeah, so good afternoon, everyone. Today I'm going to be presenting a comparative study of dark patterns across three web and mobile modalities with the goal of learning how these patterns might differ across versions of the same web surfaces and what kind of differences might be revealed by this study. And this paper was published in CSCW last year and is joint work with Amogh

Pradeep, David Choffnes, Woodrow Hartzog, Christo Wilson, and is also supported by GoogleAspire and NSF Frontiers Grants. Next slide please.

So, to start off, what are dark patterns? So generally speaking, they're often defined as designs that are used to trick or manipulate people towards unintended actions or situations. And chances are if you've been on the internet in recent years, you've definitely encountered one before and then in recent months they've actually garnered a lot of considerable attention for the ways that they interfere with your digital experience. And a lot of this attention focuses on privacy issues for the highly tracked world that we live in, especially when it comes to things like extracting user consent, like the purpose of this panel and Hana's work, but also for things like cookies registration or data collection.

And the example here primarily uses the preselection dark pattern, which is denoted in red on the left, which selects an option by default before the user has a chance to decide or actually click on the checkbox and make an affirmative action themselves. But this example here from Sally Beauty's web mobile app also uses a confusing consent regime that groups the consent to important items like the terms and conditions and privacy policy, which are noted in the blue line on the bottom right caption with unrelated and optional things like email subscriptions and reward program subscriptions, which are noted in purple. And these force all users to agree to these tertiary options that they may not have wanted but only to create a simple shopping account and to actually consent to the terms that they need it to. So, designs like these make it difficult to understand what you're signing up for. They don't give you real choice and they trap you in the decisions that you make.

But beyond consent interactions, dark patterns are also employed in diverse ways to extract other things like attention, data, and money from you by interfering with your autonomy. So, in academic circles, researchers have built really great taxonomies of dark patterns that help explain how they work, what makes them dark, and have also investigated dark patterns in mobile apps and e-commerce websites and a lot of new other services coming out in recent scholarship. Next slide please.

So, the prior investigations before this work that we conducted looked at apps and websites in isolation, which is awesome, but also leaves us with open questions on how dark patterns are used by the services that can offer both app and web modalities of the same service. So, a lot of the things that we use online today can be accessed virtually anywhere anytime through a variety of platforms and device types. For example, with smartphones, tablets, computers and more. And these platforms constrain the user interactions through different affordances capabilities, design norms, and even things like the size of the screen that you're interacting with. So, if the user experience differs between these interface types, so might the experience of darkness that they encounter. And if the experience of darkness that people encounter is unequal across different versions of a service, then the versions that have relatively fewer dark patterns might incorrectly imply that the service is actually benign overall.

So, following this concern, our work then asked these two questions on dark pattern consistency. We wanted to learn whether dark patterns differ in quantity, which is the number of observed patterns per service or if they also differed per dark pattern as well as qualitatively, which means the types of dark patterns that we might observe in each modality. And in short, we find yes, dark patterns do differ across versions of a service, both when we look at the count and the types of dark patterns that we found across these. And we ultimately found that the dark pattern count was frequently higher in apps as opposed to websites both when you look within those services or across the types of dark pattern. Next slide please.

So, to arrive at this answer, we built upon prior manual content analysis methods for observing live online services and we examined 105 popular web services including some of the logos seen here, and they span 29 categories from the Google Play store for apps. We then recorded five to six minutes of interaction for each service in each of the mobile app, mobile browser, and desktop browser modalities

for those services. And after that we reviewed the recordings and manually classified instances of dark patterns using a code book of 50 cases that we mapped to the prior taxonomies that had already been designed. So, these recordings took over 30 human hours to complete and reviewing these took approximately three times as long to label and validate between two authors. Next slide please.

So earlier I said that dark patterns do differ within services, but let's explore how they differ. Across all services, we learned that each contained at least one dark pattern with the average and median services containing around seven to eight. So, of these, the averages and medians were closer to seven dark patterns for either browser, but closer to eight dark patterns for apps. And we also found that the disparity in the number of unique dark patterns per service could also vary up to 10 dark patterns between the different modalities of their provision. So, for a broader look at these disparities, the chart you see here, which is the cumulative distributive function of the Jacquard coefficient between pairs of modalities. I know that's like a lot of words soup, but essentially this coefficient represents the numbers of overlapping patterns that are divided by the union of all patterns found between the two modalities, thus a higher coefficient denotes higher similarity.

And from this we learned that the intersection or the overlap of shared dark patterns between modalities is more similar between browsers but not between either of the browsers and the app, which indicates that the app versions are often inconsistent with the other modalities. So, from this we conclude that modalities do matter when we take a look at online services. A version-agnostic view of dark patterns doesn't account for different experiences of the same service. For example, for an app user versus a desktop user, which can contain different dark pattern counts and dark pattern types. Next slide please.

So, with that in mind we then turn to a different question which is how the dark patterns differ in representation across the three modalities. And overall, again, we find a skew towards more dark patterns within apps, with more than half of the dark patterns in the code book appearing more frequently in apps when you consider each of the patterns individually.

So then to better understand what kinds of dark patterns these might be, we grouped our code book of dark patterns to different context categories, which are things based on different interaction areas like the vendor or the user objectives with the service or user interface, the types of provided features or dark pattern purposes. And we then calculated the number of services that contained at least one dark pattern per category. And while a few of these patterns are fairly consistent between modalities as noted by the vertical narrow ovals, other substantially favor apps, which is noted by the wider horizontal ovals, which we find interesting particularly for the engagement and location context when you consider that smartphones provide opportunities to track and engage users anytime, anywhere. It's also concerning for us that the context for first using a service or for trying to leave it, which is when you register, or when you are trying to delete your account, can also skew it towards apps as these types of users can trap people into services with a disadvantaged data collection practices. Next slide.

So now that we've established that modalities do matter and that apps often have more dark patterns, what does that mean? So, let's look at an example. Here, in this instance, we wanted to know whether services would force users to manage their similar settings individually, which is taking each of the toggles and manually having to handle them, which takes more time and effort, or whether services made this easier for users by providing bulk or choose all, toggle all options. And this example from Vrbo, which is similar to Airbnb as a booking service, we looked at notification settings in each modality and we found the left screenshot in the app and the right in the mobile browser.

Now, unlike the app, the mobile browser offered to select all bulk option for some of these settings, which tells us at least from a design perspective, that the service is capable of implementing these sort of interfaces and designs like these do benefit the user by making tasks more efficient and less time

consuming. Because you can imagine if you have much more than three or dozens of different settings, trying to toggle each one of those would be a lot of time and labor. So, we kind of find it odd when beneficial features like this aren't universally provided across different versions of a service.

But when we take a closer look at this, we found more cause for concern. The mobile browser did offer a bulk toggle like we said, but this was in part because it actually offered users two types of notifications settings. In comparison, the app only provided one, which was push notifications, and yet we created accounts using an email in both modalities, which kind of makes us wonder why the app didn't offer email settings to users like the mobile browsers did. Next slide.

So dark patterns make for a suboptimal digital experience overall when they make tasks like preference management or consent overly burdensome for everyone. But the Vrbo example illustrates that people can experience different outcomes depending on what a modality allows them to access or how a modality presents those options in the first place. So, for example, when options for things like settings or account deletions are missing from an app, then an app only user will have no way of knowing that the controls that they should have access to are available on other modalities and within the app that they use, they will have no way of exercising that control. So, these kind of inconsistencies across modalities can unfairly exploit users who have limited access to other versions of those services. And that becomes even more concerning when you account for the socio-economic disparities in device ownership and home internet access.

And there's research that indicates that lower income users often rely more heavily on smartphones, and they own less devices than higher income users, which raises pretty concerning questions of fairness for a service that offers both app and browser versions of their service.

So, what then? To bring back an earlier point, we know that interfaces differ. We all know that the screen size of a phone is very different from a screen size on a computer, but we're not so convinced that the interface should result in designs like the Vrbo example where really helpful interface elements and features are completely missing in one modality, especially for features that can help people exercise control or autonomy over their accounts. So, user experiences of agency and autonomy should remain consistent no matter where or how a user accesses a service in order to keep online experiences fair for everyone. Next slide please.

So, in our paper, what we explore and discuss our ways for different parties to help equalize these experiences across modalities. For example, designers could try advocating for standardized user agency and autonomy features across modalities, whether that's within their own design teams or otherwise. And this can help reduce the risk of delivering on equal outcomes and unequal experiences for different users. Researchers can also investigate how these different interfaces and different modalities, which can include things like various platforms as well as other device types like Internet of Things or VR, might facilitate dark patterns and also further inspect how dark patterns can impact those different populations who may be served by different parts of a service. And lastly, policy-makers like here today, can gain a more complete understanding of a services compliance beyond what was said in the terms of use or a privacy policy, which may help improve audits and enforcement as well as future regulation when understanding that a service that seems to be compliant if you only look at one modality may actually have issues when you look in others. Next slide please.

So, a lot more work needs to be done to fight dark patterns, but there's a lot of really great effort underway already. So, for example, last year Apple announced that iOS developers do have to provide and are mandated to provide a way for users to initiate account deletion within an app. So even if those apps or those service providers had the deletion availability in desktop, it now has to be in the app, as well.

But we also have a lot of really cool things like from the FTC with the action against Credit Karma and other dark patterns cases, but also the DSA Digital Services Act in the EU, which was approved, and I think enforced or began to be enforced around last month at last released or so. Similarly, you have things like the proposed decisions to the CCPA and the implemented articles on dark patterns in the CPA and CPR, which do require that the number of steps to opt-out of certain choices be equal or less than those to opt-in, which we hope will result in services ensuring that opting-out isn't more difficult or even impossible in one modality over another. Next slide please.

So, in sum, we performed a comparative analysis of three versions of popular online services, and we ultimately found that the dark pattern experiences are not equal between the three and often skewed towards apps. And we additionally stress the importance of equalizing user experiences of autonomy and agency across versions of a service. And we explore implications for different stakeholders. For greater detail, we encourage you to check out our paper or reach out to our team. Thanks so much.

Min Hee Kim:

Great. Thank you, Johanna. Yeah, I think it's great to point this out because it's probably easy to forget that what you see on a website when you have a full desktop screen might actually be very different from what you see on your app or even using a mobile web browser.

So, we're going to move next to our Q and A and we're going to explore these and some of the other aspects of interfaces and dark patterns. So, I'll turn it over to Gorana.

Gorana Neskovic:

Thank you. And thank you to both of our presenters. We'll turn to Hana first. Could you elaborate more on alternative consent mechanisms?

Hana Habib:

Sure. Yeah, I think I kind of highlighted that this might be a potential solution to some of the burden that users face with cookie consent mechanisms. Obviously, they're going to have their limitations as well. But essentially what these browser-based mechanisms can do is allow people to set their preferences just once and have those preferences be communicated to a website automatically. So, such mechanisms might include a browser setting, like the proposed global privacy control or the proposed scheme by the UK's ICO, which would be that browser setting where people just set once, and that preference is communicated to websites.

There are also browser extensions that are already out there and available for consumers to use that do something similar. So, they allow people to select their preferences once related to different types of cookie usage and actually block these popups from appearing to begin with. And there are also other types of browser extensions that directly manage cookies rather than relying on websites to kind of handle the cookie consent decision made through the consent interface. So, these extensions categorize cookies in a certain way, and if there are cookies that come from a website that the user doesn't want to be on the browser, it just deletes those or doesn't prevent those from being set. Like I mentioned , so these automated mechanisms also have their own drawbacks. So well one of that pricing decisions can be contextual, so there really needs to be ways for people to set exceptions or easily communicate a decision that wasn't pre-configured, or maybe these cookie consent interfaces might play a role in the future.

Gorana Neskovic:

Thank you. We're going to touch on the browser controls a little bit later in our discussion. But first, could you summarize your recommendations for the design of cookie consent notices that avoid dark patterns?

Hana Habib:

Sure. So similar to our best-practices variants, to avoid dark patterns, it'd be really important for consent interfaces to offer in-line options with equal parts for both privacy protective options and those that enable additional features. We note this in our paper, it's also important to consider usability beyond dark patterns. So, for example, designs might consider adding easily accessible definitions for cookie categories to help people comprehend what different terms mean. And websites might also consider a persistent mechanism like that cookie preferences online to help people later change their decisions and navigate some of the maybe difficulties that come with going through settings later.

And I think there also needs to be considerations for how prominent the interface is to achieve actual informed consent. So, in our study we found that people might just ignore the banner or consent interface if they're allowed to, but obviously forcing them to make a decision through a consent wall comes at a high cost, especially for frequently made decisions like cookie consent. And so, this is where automated mechanisms might help with things.

Min Hee Kim:

Yeah, thanks Hana. I think in talking about those difficulties, we'll turn to Johanna to talk about harms generally. So how do you and your team think about it? Are all dark pattern harms, privacy harms in some way? And what are some ways to mitigate the resulting privacy harms from dark patterns, whether they're direct or indirect harms?

Johanna Gunawan:

Yeah, so I think it's really interesting, right? Because while our work was more about dark patterns generally, when you think about it for the privacy context for today, if you think about how smartphones with what they have access to by means of their sensors, by means of how often they're going to be with you on your person, where they can go, ultra portability with people, virtually anything that could be done on a smartphone or an app. In this, any data collected could be used against the user towards those kind of *de minimus* individual harms or possibly even more nefarious harms like we've seen people who have had their location data leaked and have been stalked by other people.

So when you have these hyper tracking sort of mechanisms and you have dark patterns that try to buy for more engagement and try to get the user to provide more data to a service, it's kind of easy to think that there's at least some sort of relationship between the increased engagement and when you think about the context of the device that might be on. So for us, it's kind of concerning when we look at it and see that apps tend to have more of those dark patterns.

But I mean naturally of course, there's a lot of work that needs to be done in terms of effectively measuring the impact of those dark patterns. I mean, in terms of how they might be a proxy risk for privacy harms, even if it's not quite so obvious from the interface itself. But I do think that it warrants a lot more research and consideration.

So in terms of mitigation, I think it's also a fascinating sort of area because if you look too closely at one specific type of design, so for example having a visual hierarchy where one button is brighter or bigger than the other, you also risk regulating the kind of designs that might actually be helpful for users. So I think for mitigation purposes, a lot of the approaches that are based on principles, that could be a good sort of start for realigning how we think about manipulation and deception and nudges online and how

a lot of those smaller decisions, smaller design decisions can build towards those sort of larger issues with manipulation. So again, when some nudges may be beneficial in some regards or at least may provide some value, it is important to take a look at the assymetry between the company's benefit and also the user's individual benefit and to closely scrutinize whether that's so asymmetrical to the point that we're actually causing harms and severely disadvantaging users.

So yeah, I think at an abstract level for mitigations, it's really important to have regulations that don't only prescribe for specific designs and specific types of technical measurements in terms of how these dark patterns are presented, but also to help more look at some of those principles to help move the industry more broadly towards responsible computing.

Min Hee Kim:

Yeah, that's great. And maybe if we can pull on that thread about apps faring worse than browsers, if you wouldn't mind talking about that a little bit more. And then in your presentation you had mentioned the considerations like different socio-economic factors and what it means when the lower income families have more access to mobile than they do to desktop. And maybe as a part of your response, you could frame it in terms of speaking about needing to switch modalities to do things like sign up or cancel subscriptions, so that'd be great.

Johanna Gunawan:

Realize I was on mute. Yeah, I think I kind of mentioned this, I touched upon it a bit in the presentation, but the context really does matter, right? And the constraints of the device, or at least the method of interaction that you're working with. When you think about the screen size for apps, you think, again, like I said, the different types of sensors that are available to smartphones and also just the kind of difference between session-based, more temporary use of desktops where you sit down, you type something on your laptop or a physical desktop and then you get up and you turn it off or you close the screen, but a lot of us are carrying our smartphones with us and so I think, when you think about the people who only have access to smartphones, because they're still quite expensive actually.

And so when that's kind of your only method of accessing very important things like mobility for jobs and a lot of the normal functions that people need to do in order to do a lot of things like banking online. When you have a always on hyper portable device and that becomes sort of a dependent device for someone where they only have access to that, you can sort see that there's going to be a lot of issues with the potential for having a really strong context and that person does, which you can see feeding into a lot of these sort of highly surveillance different types of data collection. For example, just because engagement data is not... Just because it seems like it's only used for marketing or behavioral advertising purposes for now, it doesn't mean that that data set can't be abused later on. And so when you have proxy interactions on these type of devices that can collect so much and that can go everywhere with us, it becomes, at least to me, particularly concerning.

But then when you're trying to switch between modalities, I think that's also fascinating because in some cases it also forces people to do certain things not always to their benefit. So sometimes if you open up things like an embedded browser page, like an app, for some reason won't let you access settings within the app itself and it either opens up the browser on your phone or opens up an embedded browser link and requires you to log in again, you can actually risk people having cross-platform tracking and data flow can profile them across those two different modalities as well. So that's something that can also be problematic. And particularly when, for example, if you have mostly web services like desktop or mobile browser services that sometimes force people to use apps, then you have the privacy implications of making them have to use that application instead, which goes into the

accessibility and equity barriers that you have when you're trying to force people to switch modalities just to exercise certain controls.

Gorana Neskovic:

Thank you, Johanna. We're coming up just, we're under four minutes left, I can't believe it. So we wanted to touch upon the business case for adoption here. What do you think might incentivize businesses to adopt some of the changes we discussed today, both the user-friendly cookie consent and the consistency across modalities? Hana, do you want to take that?

Hana Habib:

Sure. Well, first, I want to highlight the obvious role that regulation has in incentives. We know that the E-privacy Directive and GDPR, which led to cookie consent interfaces to begin with, and now the CCPA and CPRA, which has introduced additional privacy choice mechanisms and explicitly banned dark patterns, so they've been important and vital in moving things forward. I think the research community has a role in identifying what these dark patterns are for specific contexts, which can aid regulators and enforcement of these rules, the ones that are existing as well as future ones.

But beyond regulation, I think there's also this business case of being known as a privacy-forward brand. We know that major companies have incorporated privacy as part of their branding, and the market and demand for privacy-enhancing technologies does seem to be growing, so I think privacy as a business case serves as a very important incentive as well.

Min Hee Kim:

Yeah, that's great, and...

Hana Habib:

Thank you.

Min Hee Kim:

Oh, thank you. Yes. And since we're coming to the close of our panel, we'll just wrap up. I'm going to turn to Johanna and just sort of ask the broad question of, what takeaways do you have for businesses and consumers when it comes to these interfaces and dark patterns?

Johanna Gunawan:

Yeah, I think it's kind of similar to what Hana was saying to the previous question, that when users are able to trust, not just for privacy purposes, but also that the vendors are not trying to manipulate them, it makes for a better user experience overall. We have great UX teams in all these companies who are trying to do their best to make great experiences, but that can also provide a sort of competitive advantage when people can trust your product, that it's going to protect your privacy and that it's going to give you the kind of experience that lets you exercise that autonomy. I think that's something that might be helpful for businesses to keep in mind when trying to think about what might incentivize them to combat dark patterns, and also for consumers to be savvy by looking for those companies that do actually take the care to implement these across their versions of their services.

Min Hee Kim:

Great. And I'll turn to Hana with the same question, please.

Hana Habib:

Yeah, so I guess my main takeaway for businesses would be to test their cookie consent interface design. As we found in our study, even small design choices make a big impact in the usability and influence the way that people use the consent interface. Our work and work by others can be used as a blueprint for conducting such evaluations. For consumers, I think it's important for them to be kind of knowledgeable about the choices and tools that are available to them and finding resources to help them navigate kind of this, the ordeal of privacy management as it currently is.

Gorana Neskovic:

Thank you so much. Thank you both. We're sadly out of time. And next up is our panel on ad tech, and I'll turn it over to Benjamin and Mike.

Ben Smith:

Hello. Welcome to panel seven on the economics of advertising technology. I'm Ben Smith, an economist in the Bureau of Economics. I will be co-moderating this panel with Mike, and I'll let him introduce himself.

Mike Sherling:

Hi, I'm Mike Sherling, an attorney in the Division of Privacy and Identity Protection.

Ben Smith:

We've assembled an excellent panel of leading experts in privacy and ad tech. We're very excited to hear about their research today. We'll hear from Alessandro Acquisti and Eduardo Schnadower of Carnegie Mellon University, and then Eric Zeng of Carnegie Mellon as well, and finally, Cristobal Cheyre of Cornell will close. Let's begin with the presentation from Alessandro and Eduardo on behavioral advertising and consumer behavior.

Alessandro Acquisti:

First, I would like to thank the FTC for this opportunity, and second, I would like to contextualize this presentation within the broader research agenda my group at CMU is working on. The empirical research on the economics of privacy has predominantly focused on highlighting the cost of privacy regulation, and unfortunately, other questions of great importance to policymaking have been paid less attention, in particular, research questions that test and vet the claims made regarding the economic benefits of targeted advertising. Some of those research questions relate to the allocation of surplus from the data economy. To the extent that economic surplus is created from the collection of data, how is the surplus allocated to different stakeholders? How much do the different stakeholders spend [inaudible] data economy? So, much of my current research and that of my colleagues focuses on these questions, and in the presentation today, as an example of this research agenda, we'll be focused on the consumer welfare effect of behavioral advertising. Next slide, please.

The study we present today originates from a simple observation. The value that consumers derive from behavioral advertising is more often posited than empirically demonstrated. What do I mean by this?

The common claim, especially from the advertising industry, is that behaviorally targeted ads offer consumers more relevant products and services. And indeed, prior research has shown that behaviorally targeted ads command higher clickthrough and conversion rates than non-targeted ones. However, in economic terms, this suggests a reduction in search costs for consumers, essentially an increase in the matching between consumer preferences and products.

But consumer utility and consumer welfare are not only affected by search costs. They're also affected by other factors such as the prices the consumers end up paying when they buy products associated with targeted ads, the quality of those products, the quality of the vendors, and so forth, and no research so far has quantified these other dimensions of consumer welfare. Therefore, no research has clarified the actual consumer welfare effect of behaviorally targeted ads. To tackle this research question, we need a counterfactual approach. We need to compare various components of consumer utility across alternative offers consumers may face, and this is what we try to test in our experiment. Next slide, please.

We designed a within-subject online experiment in which we compare along multiple dimensions vendors and products in targeted ads to competitor vendors and products in search results as well as to random products. Essentially, we found that vendors in targeted ads in the experiment tended to be of lower quality and products in targeted ads tended to have higher prices compared to vendors and products in search results. Eduardo Schnadower will be providing the details. Next slide, please. And I pass the stage to Eduardo.

Eduardo Schnadower:

Our experimental design consisted of two stages. In the first stage, participants were asked to visit randomly-selected websites and collect URLs for ads that they saw on those websites. Then, on an intermediate stage, we used scripts to visit those URLs and search for similar competing products. Also, RAs helped us collect objective product and vendor information from those URLs. In the second stage, participants were presented with sets of products in randomized order, and we captured subjective metrics in a within-subject survey. Next slide, please.

We compare vendors and products from ads that were seen by participants, vendors and products that we obtained from search results, and vendors and products that were seen by other participants and were randomly selected. Our objective metrics include vendor quality from SiteJabber and the Better Business Bureau and prices for identical products sold by multiple vendors. Our subjective metrics include purchase intentions, perceived quality, price, fairness, relevance, and familiarity with brand and vendor. Next slide, please.

We collected 487 participants from Prolific Academic, and in total, they provided 1,169 ad URLs, and we observed high concentration of vendors in both ads and search results. But as you can see in the graphs on the right, ads show less popular vendors. We see that in search results, the results are dominated by companies like Amazon, Walmart, Target, and we see less popular vendors in the ads. Next slide, please.

In terms of purchase intentions, we see that the purchase intentions were low in general. We used Likert scales in the range of one to seven, and we observed that purchase intentions were well below four. Four is the neutral point, so they were low in general, but we found that they were similar between ad and search and significantly lower for random products. In terms of price fairness, we observed that prices were, in general, well-received by participants, but they were significantly better received for the search results, and they were similar between ads and random products. Next slide, please.

In terms of relevance, we see that the products were not very relevant to participants. They were a little bit below to the neutral point, but we see that the random products were significantly less relevant to

participants than ad or search. In terms of perceived quality, we observed that in general, participants considered the products shown to have good quality, but the product quality was slightly higher for the ads than search or random. Next slide, please.

In terms of familiarity, we observed that ads have a significantly lower familiarity of vendor than search results, and also, random has a lower familiarity than both. This is because the search results came from these big companies that almost everybody knows, and in terms of brand familiarity, we see that the brands were not very familiar to participants, but they were significantly more familiar in the ads than in the other conditions. Next slide, please.

In terms of objective measures, we considered vendor quality and prices. From the Better Business Bureau ratings, we observe that vendors in ads had much higher frequency of F ratings, which is the worst possible rating that the Better Business Bureau can give to a business, than what search results have. If you see the table on the right, you also see that the highest possible rating was more frequent in search than in ads. Also, the SiteJabber rating was significantly less in ads than in search results.

Finally, in terms of prices, we found that if you do a search for a product that is identical to what you see on an app, to find a different vendor that sells that same product, your expected savings from doing that search is 11%. That's conditional on the product having price dispersion, so if a product, for example, an iPhone, has the exact same price for every vendor, we did not consider that product here, because it's a product that has a fixed price. Next slide, and I pass back to Alessandro.

Alessandro Acquisti:

Thank you, Eduardo. To summarize, we found a high concentration of vendors in both ads and search results, but ads do show less popular vendors, and also, ads are more relevant than the products selected at random. So, in these regards, behavior targeted ads are useful. However, we also found that ads tend to come on average from lower quality vendors and tend to be associated on average with higher prices, which decreases consumer utility. Because these results were of note, we decided to conduct a replication study, which was ongoing at the time we submitted our manuscript to Privacy Con. The replication study is now completed, and I can share with you that it replicated all the results presented in the prior slides, with an aggregate sample size across the first study and the replication study of close to 1,000 participants. We also have a latent utility analysis ongoing.

So, what can we learn from this? As I mentioned at the start, the economic value theory of targeted ads is more often posited than empirically estimated, and no research so far has quantified those dimensions of consumer welfare other than search cost. We designed and conducted a study which relies on the counterfactual approach, testing for the comparative impact of ads to search to random products. And if there is a term, perhaps, to summarize our results, it's that the economic impact of behavior advertising is more nuanced perhaps than we may have believed before. Thank you for your attention.

Ben Smith:

Next is Eric Zeng.

Eric Zeng:

Hi. Thank you for having me. My name is Eric Zeng, I'm a postdoc at Carnegie Mellon University, and I'm here to present my paper, What Factors Affect Targeting and Bids in Online Advertising? A Field Measurement Study. And this is work that I did during my PhD at the University of Washington in the Security and Privacy Lab in collaboration with Rachel McAmis, Professor Tadayoshi Kohno, and my advisor, Professor Franziska Roesner.

The online advertising ecosystem is really opaque. There's just a lot we don't know about how targeted advertising is used in practice, like how prevalent it is, who's being targeted by what, and how effective all this targeting is, and that makes it hard to make data driven decisions about some of the big policy questions surrounding privacy and advertising. For example, should there be new regulations to limit the amount of data that companies can collect for targeting ads? Should we remove or replace third-party cookies? And how would these measures hurt advertisers? Since we don't have a clear picture of targeting as it exists now, it's hard to understand what effect these changes would have. Next slide, please.

So, to shed light on what is happening in this ecosystem, we conducted a field measurement study to broadly describe the targeting that end users are experiencing on the web. In particular, we wanted to investigate three questions. So first, how prevalent are behavioral targeting and contextual targeting, where behavioral targeting is based on the interests and characteristics inferred from users' browsing behavior, and contextual targeting is the targeting based on the website the ad appears on?

Basically, we wanted to know how much of the targeting on the web actually requires all of this privacy-invasive data collection from users. Second, how do ads differ across demographic groups due to behavioral targeting? We know that people see different ads and that discriminatory targeting outcomes can happen, so what do these demographic trends actually look like in the wild? And third, how much do advertisers bid to place their ads in ad auctions, and what factors affect these bids? We're interested in these bid values because the amount that advertisers bid can tell us about which signal is about people they find valuable. The goal of measuring all these things is not to argue for a particular course of action, but just to provide a broad view of what targeted advertising looks like and to point to areas we might want to investigate further. Next slide, please.

Next, I'll give an overview of how we designed our study. Our overall goal was to measure end users' experience of targeted advertising. First, that meant collecting data from real users rather than using a web crawler. For many reasons, it's difficult to run crawlers that accurately reflect real users' browsing behavior, so for this study, we built a browser extension that people participating in our study could install, and this extension would send us screenshots of the ads they saw and the winning bid values of those ads. This means this data we collected reflects the targeting of real people.

We also sought to capture a demographically representative sample of people in the US so that we could analyze any differences across demographic groups like age, gender, and ethnicity. And lastly, to make the study more controlled, we scoped our data collection to a fixed set of websites so that the differences that we saw could be attributed to differences in people's advertising profiles rather than the differences in the set of websites that each person decided to visit during the study. Next slide, please.

Next, I'll talk a bit about the procedure of this study. First, the study was approved by our university's IRB. We recruited a demographically representative sample of 286 participants from Prolific, an online research panel. The participants in this study first installed our browser extension and then visited a fixed list of 10 websites that we selected, and these websites contained a variety of topics, such as news, weather, and technology. Next slide, please.

Moving on to the results, first, we found that contextual targeting or targeting by website is clearly being used on some websites but not others. For several sites in our dataset, the majority of the ads were closely related to the topic of the website. For example, on businessinsider.com, a business and financial news website, we observed that 60% of the ads were either for business products, careers, or credit cards, and on phonearena.com, a smartphone review site, about half the ads were either for electronics like smartphones or for mobile phone service.

However, on other sites, the categories of ads were more evenly distributed, and the top categories were less topically relevant. For example, on weather.com, some of the top categories included medications and food and drink. I think this shows that there's parts of the web where contextual targeting makes sense for advertisers, and they do so more heavily, but there's other sites where this context alone is not a good enough of a signal for advertisers to target, and so the ads there are more likely to be targeted behaviorally. Next slide, please.

We also saw evidence of behavioral targeting at the individual level. Our data showed that different people were seeing different kinds of ads despite visiting the same websites. We can visualize this in a Lorenz curve, which is typically used to show distributional inequality like income inequality, but here, I'll use it to show how ads of a certain category are unequally distributed. Each line in this chart shows the proportion of ads in a particular category seen by the bottom n% of people. So, if everyone saw the same number of ads in a category equally, then the line would be straight, and then the deeper the curve, the more disproportionately the ads are shown to a small fraction of people.

We can see that, for example, electronics ads were relatively equally distributed, but other ads were shown to a smaller set of people. For example, only half of participants saw health insurance ads, and 34% of health insurance ads were seen by the top 5% of people. I'll just note that this isn't a general finding about these particular categories of ads, because this data is biased by that list of 10 websites where we collected data on, and those are pretty heavy on technology-related topics. But it does show that there is behavioral targeting that seems to be happening and causing different people to see different kinds of ads. Next slide, please.

Surprisingly, we didn't find large differences in the types of ads seen across demographic groups like gender, age, and ethnicity. We saw some individual examples of differences across demographics, like women received significantly more ads for apparel and beauty products, and 45 to 54-year-olds saw more ads for jewelry. However, the overall magnitudes of the differences are relatively small, and we didn't see differences across most ad categories. Overall, only 9% to 16% of the ad categories were over or underrepresented across these three demographic groups.

I think there's a few explanations for these relatively small differences. First, we were looking at fairly broad categories of ads that might hide some differences that we would see if we looked at it more granularly. For example, we didn't break out apparel ads between men and women's apparel, but if we did, we might see more obvious differences by gender. But second, it might be the case that advertisers might not be targeting so much directly on demographics, but more on interests, which in some cases might correlate with demographics, some cases might not. And lastly, many ad platforms no longer allow advertisers to target by ethnicity, so another possible explanation there. Next slide, please.

Another type of data we collected is related to how much advertisers bid to place their ads. This measures how valuable a particular impression of the ad was worth to the advertiser. On average, we found that the median winning bid was $4.16 per 1,000 impressions, or CPM in the industry. Again, we saw there were differences across individuals, whose average bids ranged from as low as $2 in CPM to over $10 in CPM. We also saw differences across websites, with some sites having a mean as high as $9.95 or as low as $2.44. However, we again just didn't find any correlation between bid values and demographics, and so basically, what this shows is that advertisers are valuing certain people and certain websites more than others. Next slide, please.

One other interesting thing we found from the bid value data is that high bid values appear to be correlated to retargeted ads, and this replicates some findings from other work in this area. A retargeted ad is a form of targeting where if you visit a website and you look at a product, let's say you're looking at some shoes, and then later you go to a different website, and you see an ad for those shoes, because that website basically remembers that you had visited them and they're targeting you with an ad that

for that product later. This is a particularly interesting form of targeting, because it is behavioral targeting, it's based on your activities, but the advertisers don't need to collect nearly as much data on your web history to send these ads.

We asked our participants to tell us which ads came from sites that they visited previously and then compared the prices of those ads to the rest of the ads. Participants reported that about 18% of the ads they saw were from sites they visited previously and that advertisers paid $1.07 per thousand impressions more to place those ads, which were likely retargeted. Interestingly, at the very tail of the distribution, we saw some very high outlier bids that were as high as $89 per 1,000 impressions, or a whole $.09 just to show a single ad. This suggests that this form of targeting is actually quite valuable to advertisers, based on how much they're willing to spend. Next slide, please.

I'll end this talk with a few potential takeaways and future directions. First, our data shows that alternatives to behavioral targeting on the web, like contextual targeting and retargeting, are both prevalent. This raises the question, what would the web look like if advertisers could only use contextual targeting or retargeting? Could they still effectively target people? Because if they could, then it suggests that we could really limit the amount of data that companies can collect for targeted advertising purposes with relatively little impact. But if it is impactful to limit this type of targeting, then maybe we need to consider compromise solutions, like the FLoC or the Topics API proposed by Google, as a sort of middle ground.

Second, we found that it was pretty difficult to investigate demographic disparities in targeting among real users, and it might have been that the differences were too small to detect from this vantage point. I want to say that that doesn't mean that there isn't discriminatory targeting happening, but it's just we might not see it outside of more narrowly-focused studies. For example, there's a bunch of great work from Northeastern University where they audited Facebook to test whether employment and housing ads were specifically delivered in a discriminatory way, and they did this by buying their own ads on Facebook. These studies are really great and illuminating, but also, these methods don't let us see how prevalent these disparities are in the wild.

Maybe that leads to my third and broader point, is that we just really need more transparency from the ad tech industry on targeting. There's a lot of open questions about these privacy tradeoffs with targeting and discrimination in targeting, and there's just a huge barrier to conducting these kinds of audits, because we just don't have the data, either on the targeting parameters being used by advertisers or any kind of general metrics on targeting outcomes. So, having this direct access to this kind of data would enable much broader and more accurate audits by researchers. And with that, thanks for having me here, and thank you for watching my talk.

Mike Sherling:

Thanks very much, Eric. Next, we will turn to Cristobal for his analysis of the GDPR's impact on content providers.

Cristobal Cheyre:

Hi. First of all, I want to thank the FTC and the organizers for putting together this panel and inviting me to be part of it. I'm going to be showing today a longitudinal analysis of the impact of GDPR on content providers. This is joint work with Vincent Lefrere, Logan Warberg, Veronica Marotta, and Alessandro Acquisti. Next slide, please.

When GDPR was enacted and implemented, there was a lot of excitement and optimism about the positive impact it could have on a consumer privacy, due to how comprehensive it was and due to the fact that it did not only apply to sites based in the EU, but to any site dealing with the data of EU

citizens. Along with all this excitement, there was also a lot of concern from the ad-supported online ecosystem, all the websites that derive most of their revenue from online advertising, because if GDPR went too far into curtailing the use of private data, this would significantly affect availability of free, online ad-supported content on the internet.

In this study, what we do is to analyze the impact of the GDPR in online ad-supported content providers, specifically on news and media websites. Just to preview what I'm going to show you, we do find that GDPR had an effect on how websites deal with consumer privacy on the choices that they offer to their users. It had an effect on reducing tracking online. However, we do not find any evidence that GDPR led to the exit of EU websites that were ad-supported. We don't find evidence of these websites shifting their business model to other monetization strategies, or that they were negatively affected in their ability to continue to create content or the quality of this content, measured as the amount of engagement they derive from social media users. Next slide, please.

It's important to think, why should we expect that GDPR could have a negative effect on online ad-supported websites? One of the things that GDPR requires is that before tracking a consumer, you have to explicitly ask for their consent to do so. We should expect that this leads to fewer users consenting to online tracking. That can lead to fewer targeted ads being available, as targeted ads are in general more valuable than non-targeted ads. This can reduce the revenues that online content providers get from advertising, and if online websites are getting less money, we should expect some of them to go out of business or to start producing less content or lower-quality content.

What we do in this study is to look at the beginning and the end of this chain of events. In the beginning, we look at whether GDPR affected the amount of tracking that exists. This is what we call the technical variables that we analyze. And at the end of the chain, we look at whether GDPR affected the quantity and the quality of the content being produced. This is what we call the downstream outcomes that we have. Next slide, please.

The analysis is based on following about 900 news and media websites. These are split roughly equally between EU and US websites. This includes both top-ranked websites and websites that are not in the top ranking. What we do is that we visit these websites using a privacy measurement framework called OpenWPM. We visit this from both US and EU IP addresses to simulate EU and US visitors, and the results I'm going to be showing today are based on 16 waves of data collection conducted between April 2018 and November 2019. The variables that we collect include, from the side of technical variables, things such as the number of first-party and third-party cookies used by websites, the length of advertising content published in each website, whether the website is blocking EU visitors, if they're using a consent mechanism, a cookie banner, or cookie walls, or if their privacy policies are claiming a legitimate interest to support their collection of private data.

In the side of downstream outcomes, we tried to collect the quantity of content by looking at the number of URLs posted, and we tried to measure the quality of content by looking at variables such as the number of page views per user that the website gets, the reach, the ranking, the page views per million internet users that the website gets, and the reactions that the content published by the website gets on social media. For example, the number of likes, the number of shares that these URLs are getting on Facebook. Next slide, please.

We expect GDPR to have an effect at both the ecosystem level and the website level. Ecosystem effects refer to a differential impact the GDPR may have on EU websites versus US websites. It is true that the GDPR applies to both websites in the EU and the US, but websites in the EU are required to comply with GDPR for all of their visitors. In comparison, websites in the US only need to comply to GDPR for their EU visitors. It's natural to expect that EU websites get many visitors from the EU. US websites to get very few visitors from the EU, so we should expect EU websites to be more affected by GDPR.

At the same time, as all websites in the EU have to comply for all their visitors, we should expect that the amount of aggregate data available regarding EU consumers is also going to decrease over time. This can make the ability to make inferences or to target advertising less precise in the EU ecosystem versus the US ecosystem, and this is another reason why we expect the EU websites to be more affected. The website-level effects are related to the impact that the specific responses adopted by the website has on downstream outcomes. Next slide, please.

The website-level responses that we're looking at is whether the website is blocking EU visitors. Some websites, especially those outside of the EU, may decide to not deal with compliance and exit the EU market completely by blocking all this results from the EU. Some websites may choose to stop using any sort of tracking websites may decide to adopt consent mechanism, display dialogues that request users consent before tracking them. Some websites can be a bit more aggressive and require users to consent to tracking before allowing them to see any of the content. And some websites can choose not to do anything in response to GDPR or to claim that they have a legitimate interest to continue doing what they were doing before without changing anything in the privacy practices. Next slide please.

So I'm not going to show all the results in the paper. I just wanted to highlight a few key results from the side of the technical variables. If we look at the number of third party cookies using by the websites, we can see a very different behavior between EU sites and US websites.

So to begin with, we observe a GDPR effect in the short term. Right after GDPR there is a reduction, the use of third-party cookies in both US sites and EU sites. But three, four months out from GDPR we observe a rebound in the number of third-party cookies being used. What is more interesting is to look more towards the long term. We observe that US websites are tracking users as much or more than they were using before when it is a US website visited by US visitor. In the case of EU websites visited from the EU, we do observe that there is a significant reduction in the number of third-party cookies being used. Another thing that's interesting to note is that both EU websites and US websites treat differently EU and US visitors. US websites are very careful when dealing with EU visitors and almost stopped tracking after GDPR. Whereas EU websites are less careful with US visitor than they are with EU visitors. Next slide please.

Another factor that is interesting to look at is the type of responses used by different websites. In the left of the slide is the responses adopted by EU websites. We see that over time, for EU websites by and large, the majority of the response is to adopt consent mechanisms. On the right is the response used by US websites. There is an increase in the use of consent mechanisms, but not as much in US as in the EU, and the most prevalent response within US websites is to not do anything or claim legitimate interest or to stop the tracking of EU visitors. Next slide please.

Looking at downstream outcomes, we try many different implication studies in the paper. This result is the simplest study that we have, which is a difference in difference analysis. We just compare our downstream outcomes before and after GDPR comparing EU versus US websites, we don't obtain any statistically significant results when making this comparison with the exception of a small effect on the number of page views per user. This effect is statistically significant but is small in magnitude. Next slide please.

So having these results in mind, it is important to consider why we don't observe a stronger effect in downstream outcomes, as we observe on technical outcomes. So the first explanation is that maybe advertising revenues did go down, but the website aggregate revenues did not. This could happen, for example, if websites shift towards other monetization strategies such as using paywalls or highlighting subscription models. We tested this, and we didn’t observe any evidence of EU websites shifting towards these other monetization strategies. Another explanation is that maybe websites see a decrease in the revenues generated by impressions, but they compensate by showing more

advertisements per page. We examines the amount of advertising intensity in webpages, and we didn't observe any increase in advertising intensity in the EU versus the US.

Another potential explanation is that maybe GDPR did not reduce the availability of user data or the ability of website to target advertising—Maybe there is as much tracking as before. So we do observe that website responses change over time. They are experimenting with different things. However, I would contend that over time or the long term, the response that is predominant is the use of consent mechanism. Someone may contend that maybe these consent mechanism are full of dark patterns that are nudging users toward consenting. However, we are also observing that these consent mechanisms are evolving towards more transparent consent mechanisms that may make predicting tracking easier. Finally, the last explanation is that maybe GDPR has reduced the availability of user data, but the ecosystem has adapted over time to continue to deliver profitable ads using less consumer data. Maybe this ecosystem can continue to operate in regimes with greater privacy protections. Next slide please.

So just to summarize from the side of website responses, we find robust evidence of responses to GDPR. We observe an initial reduction of third-party cookies and a rebound over time. However, we do observe that for EU websites visited from the EU, there is a significant decrease in the amount of tracking. From the side of down speed outcomes, the only statistically significant result we find is a reduction in the number of page views per user. While this is statistically significant, the magnitude of the effect is relatively small. And what I would emphasize is that we don't observe any impact of GDPR on the ability of EU websites to survive. We don't observe them changing monetization strategies and, when compared to US websites, they are still able to provide as much new content. Their ranking hasn't changed, and they are generating as much social media engagement. So to a great degree, it seems that EU news and media websites have continued to thrive under GDPR. Thank you.

Ben Smith:

Okay, great. Thank you to our panelists for those excellent presentations. We're going to start our Q and A session now. And this first question is for the panel generally, but I'd like to first direct it to Alessandro. So Alessandro, can you discuss the traditional understanding of the costs and benefits to consumers of behavioral advertising, and does your research change this understanding at all?

Alessandro Acquisti:

Thank you. I guess that the perception of the costs and benefits is perhaps that behavior advertising may be perhaps problematic from a privacy perspective, but from an economic perspective is advantageous for consumers. And perhaps this is a little bit of a strawman, but I'm kind of summarizing what I seem to see as a main claim in the debate around behavioral advertising. Privacy is a problem but economically it's good. This is an argument that theoretically is legitimate and valid. The problem is that as I try to espouse in our presentation, is that there is little empirical validation. So by this I don't mean that the argument is false. I mean that we all have to make a better effort trying to vet these kinds of claims. So essentially what the study Eduardo and I presented tries to do is to affect or change the frame of the debate specifically by putting attention on the economic claims of benefits so that we can actually estimate them.

So, tying our study to that of Eduardo, the two main channels for which behavioral advertising is claim to be economically advantageous to consumers is a direct channel and an indirect channel. The direct channel is the channel for which behavioral advertising increases directly consumer welfare by presenting better offers, essentially. And like I said, we believe that there is plenty of evidence that behavior advertising reduces search costs and increases matching and that's good, that's most likely consumer welfare enhancing. But there is this lack of research on the other components of consumer

welfare that may be directly affected by behaviorally targeted ads. In this regard, yes, my hope is that the research we present today helps by contributing to a broader debate over these direct claims of direct economic benefits of behavior advertising. As for the indirect channel, well that's essentially what Christopher was talking about. The notion that behavior advertising allows content providers to keep providing high quality free content essentially because behavior advertising increases publishers revenues and therefore allows these provisional high quality free content and again, on theoretical grounds these legitimate argument.

But the corollary of that is that any regulation that may somehow negatively affect consumer tracking and targeting would therefore also negatively affect the content provision, and therefore would negatively affect consumer welfare through this indirect channel. And as Cristobal has shown, this is not what we find in the data coming from a longitudinal study of GDPR. Now certainly one could say that GDPR enforcement has been so far spotty at best, and I would not disagree with that. Nevertheless, what Cristobal has shown is that there are no downstream negative effects on content provision from such a significant piece of regulation. And again, this to me is of note because he adds some nuance to the conversation regarding the economic benefits of online advertising.

Ben Smith:

Okay, great. Would anybody else like to add on that one?

Eduardo Schnadower:

So I think that in our research we found several issues that haven't been addressed before in behavioral advertising, like prices and quality. And the fact that we see vendors of low quality appearing in ads, it may be one of the ingredients that people use when they decide whether to use ad blockers or not. If you see not only that the ads know you and the ads interrupt you, but also that the ads give you something that has low value to you, that may be a motivator for installing ad blockers. And this in turn could in the long run cause this constant war of ad blockers versus content providers trying to prevent ad blockers from being used. And that may not be good in the future for the free internet economy as it is sometimes called.

Ben Smith:

Interesting. Thank you. So I'll move to the next question also directed to the panel. And it's really an observation that most privacy regulation research seems to focus on the cost of regulation as Alessandro alluded to rather than the benefits. And we are wondering if there are any promising methods to measure the benefits or if they're just inherently more difficult. And if so, why would that be? And maybe we'll start with Alessandro again, if you don't mind.

Alessandro Acquisti:

Yeah, yeah, sure. I could start offering some thoughts, right. It's a very good question, and we don't precisely know why it is the case, but I do agree with the point that the empirical privacy economics literature, it's mainly focused on the cost of regulation. Whereas, and there is an interesting disconnect here, if you look at the theoretical economic privacy literature, it is much more nuanced. It's much more nuanced because it shows these very context-dependent trade off costs and benefits of data protection and data sharing. In some cases, data protection could be welfare increasing, and in some cases it could be welfare decreasing, whereas the empirical effort is more on the welfare decreasing scenario. And one could say is that that's because most of the times regulation creates some economic unintended consequences, and maybe on theoretical grounds that's a legitimate argument. I believe there is also an

element of self-selection that when regulation is enacted, it's easy to look for short term tangible metrics such as changes in revenues of intermediaries or changes in clickthrough or conversion rates for merchants.

But it's much, much more difficult to look for the downstream implications. For instance, we know that there are many diverse types of privacy harm the consumers may face when the data is compromised or misused. In fact, I argue in a piece that the problem of consumer privacy from a cost perspective is not that there are few harms—it is that there are too many, and they come in all different heterogeneous forms in terms of probability and magnitude. And this diversity creates a sort of aggregation problem. It is difficult to capture the aggregate cost of the lack of regulation, and therefore the benefits of regulation, precisely because the costs can be so diverse. They go from price discrimination to identity theft, from stigma to other forms of discrimination, and so on and so forth. So this creates an aggregation problem, which to me is one of the biggest challenges that we face when we think about where the empirical economics of privacy can go in the next years.

Mike Sherling:

That's great, Alessandro, very interesting, and I think fruitful avenues for research. I wanted to direct the next question to Eric. Certainly, if you have any thoughts along the lines of the previous question, feel free to expound. But I was also wondering if you can describe the challenges to understanding and researching the bid auction process that undergirds the targeted advertising ecosystem. And if there are any ways or suggestions that you have for researchers and policy makers to get more insight into the ecosystem.

Eric Zeng:

Yeah, sure. I'll start by just adding on a little bit to the previous question on the benefits of privacy regulation. So most of my research is on the intersection of human computer interaction and security and privacy. And so I think there's also just a lot of benefits to consumers directly that are also sort of dispersed and hard to measure. If we had a world where there were more regulations, more privacy, people may be more comfortable using certain online services, they might find them more trustworthy. And so that could have benefits both for the consumers but also for the websites, for the apps or whatever. But these could be relatively small and hard to measure, but I think it's an alternative world worth imagining.

And then, so moving on to the question about researching the ad auctions, bid prices, and winning bid amounts. I think it's a potentially very interesting signal we can use to understand what things advertisers find valuable and are interested in. It's not always clear why it is that an advertiser will pay more for one thing or another, but I think there could be a lot of fruitful experiments where we could try manipulating things to try to simulate particular online behaviors, like interest in a particular topic or something like that. And seeing if that causes basically a spike in these prices and advertiser interests. And I think actually earlier in privacy, one of my UW colleagues, Omar actually used this technique for the Amazon Echo measurements.

I think if we could reliably measure these winning bid values, I think that could be really interesting. But again, this is one of the things where there's only small set of websites where this actually works using this technique. And so just again, another area where we could use more transparency so that we could actually freely conduct these kinds of audits rather than just working really hard to get a tiniest little window into what's going on.

Mike Sherling:

Certainly. So we only have a few minutes left and I wanted to turn to Cristobal and just to see if you can comment on how GDPR compliance has changed over time. You touched on this in your research, but there have been new developments since you reported or collected that data. Can you comment on any of the new decisions regarding cookie consent and potentially how the CCPA and CPRA and other US state laws are impacting the consent and content of websites?

Cristobal Cheyre:

Now. Thank you. That's a very interesting question. Because the GDPR was not a one-off impact—It's something that has continued to evolve over time. The results that I show today are based on a longer period than what has been done in most GDPR studies that I have seen. But it's important to continue to see how it has evolved, particularly because after GDPR there have been many events that have changed how websites behave. So for example, data protection authorities have issue notices clarifying how GDPR should be implemented. So when GDPR was enacted and became effective, there was a lot of discussion on exactly how it should be implemented and what it meant, because the text was written in such a way that needed a lot of interpretation. Data protection authorities have issue clarifications that have clarified what a website should do.

So using the data that we have collected, we have studied how these events impact website responses, and we do observe that when these events are enacted, the responses of websites change. Besides the actions taken by data protection authorities, also industry organizations have been providing their own tools to comply with GDPR. One that had a very significant impact was when the IAB released their transparency and consent framework that led to a great increase in adoption of consent mechanisms. Then that transparency and consent framework was invalidated, and now they are releasing a new transparency consent framework. And it would be interesting to analyze going forward, what's going to happen when these new frameworks implemented. How it will change the number of websites using consent mechanisms. I don't think it's going to change as much because many websites are already using them, but more importantly, how it's going to change the degree to which users consent to tracking. I think those are all things that are very important to continue to consider.

Mike Sherling:

Great. Yeah, absolutely. So I think we are wrapping up, but I will open the floor to one last question, which is looking to the future, in which area do you see the greatest need for more privacy research in terms of the ad tech ecosystem and kind of social or economic cost and benefits? And Eduardo, I'll put you on the spot to see if you have anything to say on that, and then open it up to everyone else.

Eduardo Schnadower:

Can you repeat the last part? I had some audio issues.

Mike Sherling:

No worries. I'm just wondering if you see an area in privacy research that has the kind of greatest need for more research in terms of the either economics or social costs and benefits.

Eduardo Schnadower:

Yeah, so I think that what we did with consumers is just the beginning because we just measured different metrics that were subjectively determined by the participants based on Likert scales. But there's plenty more to do to measure this more precisely. And maybe things like trying to do big field experimenting, which we can observe consumer behavior rather than just subjective perceptions would

be also a great avenue to be able to measure how consumer welfare is affected by behavioral advertising.

Mike Sherling:

Great. Anybody else?

Alessandro Acquisti:

In addition to research on vetting what I call the benefit allocation of data—how the different stakeholders share the benefits of data collection—I'm a strong proponent of privacy enhancing technologies, anything from differential privacy to homomorphic encryption, and I strongly welcome the very recent increase finally in economics of interest in this area. Computer scientists have led the way, have been ahead of us for 15 years, and finally we started catching up. That's good. We should do more research on how privacy enhancing technologies may be able to protect privacy while allowing data analytics, and what change, if any, to the costs and benefits allocation.

Cristobal Cheyre:

Yeah, I just wanted to add that something that I find very exciting going forward is moving away from the focus on the cost of regulation towards trying to measure the benefits of regulation. As Alessandro said a few minutes ago, it's very difficult to do, but I'm really excited by economists working with computer scientists to implement different field experiments to try to get into these issues. Things similar to what Eduardo showed today or what Eric showed today. I think those are pretty exciting things to look forward to.

Mike Sherling:

Great. Well thank you everybody. This was a stimulating and excellent discussion, and now I will turn it over to Jamie Hine for closing remarks.

Jamie Hine:

So in the next few days, we will make the transcripts available on the event page at ftc.gov. We'll also update any of the papers and provide some of the presentations that you've seen today. For example, Alessandro and Eduardo's paper with the updated data is already available on the website. You can check that out. I just want to say thank you to everyone today. I mean the program really wouldn't be possible without the great work of all the people at the FTC and Open Exchange who helped us do the livecast today. That includes attorneys, economists, technologists, folks in our Division of Consumer and Business Education, our Office of Public Affairs, who did a phenomenal job live tweeting today and our FTC Events team who helps make all the logistics work well. A very special thank you for me to call out two people, Caelan Conant and Molly Smith.

They are the paralegals that have sort of done the day-to-day work to make all of this work so well. So thank you to both of them. Now that we're finished with PrivacyCon 22, it means that PrivacyCon 2023 starts tomorrow, and that means that we're looking for new ideas and new topics. Alessandro has already called out Privacy Enhancing Technologies. It sounds like we're going to have a panel there, but if you have any ideas, please send those to us. We have an email address, it's privacycon@ftc.gov. We're happy to hear any ideas about this year's event or ideas for next year's event. And until next year, thank you so much. Take care.