JAMIE HINE: Good morning. On behalf of my colleagues at the Federal Trade Commission, I am happy to welcome you to our 6th annual PrivacyCon. My name is Jamie Hine. I'm in the Division of-- an attorney in the Division of Privacy and Identity Protection.

And along with my co-presenter, Lerone Banks, we're happy to bring you PrivacyCon for 6th year. A few details before we get started. We're happy to have you join the webcast. Our agenda is available on the PrivacyCon page with links to all the presented research as well as the biographies of all the moderators and presenters today.

There's also a webcast link on ftc.gov page as well as the event page if you happen to get disconnected where you're joining us sometime in the afternoon. Following PrivacyCon, we will make all the presentations available online. Usually takes about 2 weeks, but we archive all of the presentations today. You'll be able to go back and watch them again.

We'll also provide a transcript that you'll be able to read along and see all of the great presentations again today. After what seems like a lifetime assumes, we all realize that technology happens. So we ask for your patience today.

We have a technology team that's here to address any issues. But if you have specific problems, feel free to email us at privacycon@ftc.gov, and we'll try and help you out as quickly as possible. We welcome questions for our audience.

PrivacyCon is a participatory event, and what that means is if you have any questions for any of our panelists today, send them to privacycon@ftc.gov. We'll have somebody watching that email address and we'll send f w
l an d

And [INAUDIBLE] already begun to assemble a team to make some of these shifts when it comes to fashioning more effective remedies for law breakers, understanding the full range of harms and sharpening our analytical approach. I'm pleased that Stephanie Nguyen has joined the agency as deputy CTO and Presidential Innovation fellows [INAUDIBLE] and Vivian Le[illegible]

So in light of this result to discuss some implicate policy implications of our work, what are technical evidence has shown is that the role that ad delivery algorithms play in ad delivery is important in that regulation needs to take this into consideration. And the questions that we would like policymakers to engage with are first is this technical evidence sufficient to enact new policies that will mandate our platforms to change how the ad delivery algorithms work? In the past, legal challenges and also civil rights audits have pushed our platforms to change, how their ad targeting works. And so we hope to see similar action in the context of ad delivery. 00E>-3Â 0 and_0 13 <

The blue group is the majority group, which we are seeing on your network models for both fair and standard models. Here, I show you the results of a training data for two protected groups with positive label. The x-axis is a privacy risk and the y-axis is training accuracy.

The triangle marks represents the results on the standard model and the circle marks represent the results of a fair model. We can see that standard model performs much better on the blue group compared with yellow group. If we choose to use fair model, the yellow group will have a better accuracy.

However, the privacy of risk for the yellow group is also increased. At the same time, it shows that fair model improves the accuracy, but leaks more information about this underrepresented group or under unprivileged group. On the next slide, let's see the trade-off between fairness and privacy.

May vary with distribution of the data. Each [INAUDIBLE] year shows the results for one setting. The x-axis shows the fairness gap of the standard model with respect to equalized odds reflecting the unfairness of the model.

The y-axis is a privacy cost for the underprivileged group. The privacy cost is measured as differences in the privacy risk between standard model fair models. We can see a whew 2B>-6.7 Abe0D>]TJ Â 1_0` 1 Tf•( )Tj /5745 <000F>-

The first is it gives agency to individuals. So assume we have customer users out there and the decision was made by a machine, and now the person is unhappy with this decision, should happen from time to time, it's impossible for this person to argue against the decision if they don't understand it. And if you have heard one of the horror stories, for example, that the decision might change if you change the spelling of your name or the decision might change if your Main Street instead of Main St in the formula. Then you really want to understand how the decision was made so you can change it.

On a larger level, agencies like the FTC want to go in and audit a model. And if you just look at the big [INAUDIBLE] model, it's really hard to audit it. So you want to be able to explain the model so you can audit it. For example, if you want to look at whether or not the model is fair.

at a òþPÐ ð taüat< C ñmŠ$ÐÙÐ ð!ÜÂ ó r1CÐ € ð“ ‘ `

DEVIN WILLIS: Thank you very much, Martin, and again to all our other panelists for those very informative presentations. I'm rea N

ÁôypÂ` 0 Td1 10>0]5.44 7/C8 B/C2_1C00]9<0.797 0 4.797 0>0]9<0.797 0 4.7971 10>0 591C0

And I think the FTC can help because like it can regulate who has access to this transparency tools and like if auditors have access and trusted auditors have access, think there's little privacy risk. The privacy risk comes from everybody having access. So if you kind of can ensure that the right people get access and more people like Basi have access, that would be good.

DEVIN WILLIS: OK. Thank you. I mean, this has been a very interesting discussion this disclosure

ZIAD OBERMEYER: Yeah, I think that's correct because it's certainly those proxies even if they're imperfect, are certainly going to give you a readout. They're going to be correlated to the real variables of interest with all the caveats, but they're not exactly right. And I think, from an optics point of view, one thing I found is that regulators genuinely just want biased algorithms not to be used. At least the ones that we've been working with, there hasn't been like a aQ a

LERONE BANKS:

ZIAD OBERMEYER:

DANIELLE ESTRADA:

NICO EBERT:

SIDDHANT ARORA: The major research contributions of our work are the following. We built machine learning classifiers to automatically extract these opt-out choices from the privacy policies. We built a browser extension to suit the opt-outs for a given website. The browser extension is now publicly available and can be downloaded from the links shown on the slide. Another

DANIwK w w я ww

In the first stage, we asked participants to provide the most commonly used email address for [INAUDIBLE] API, followed by questions about several properties of email address such as its frequency and purpose of use. In case the participant's email was not tied to any breaches, the participants were given the opportunity to enter another email address, which they believed to be more likely to be involved in breaches. In the second phase, all participants that were affected by at least one breach represented up to three specific breaches from the full set returned by [INAUDIBLE]. For each breach then, we collected data relating to our participants awareness of the individual breach before our study, their perception of causes and impacts of being impacted, the emotional reactions and if they've done or intend to do anything in response.

And in the end we collected the participants demographics and showed them the complete list of known breaches that included the email address to ensure that they're aware of all the risks. And additionally, we provided resources to help participants in taking action and dealing with the potential aftermath of the breaches we showed them. Next slide, please.

So using the data from the survey, we aim to answer five research questions namely, the factors that influence the likelihood of an e-mail addresses' exposure to data breaches, the participants' perception of causes and impacts affected by data breaches, the awareness of the data breaches, the emotional reactions, and the behavioral responses to the data breaches. Next please. So in this talk, we will only-- I will only present results regarding four of these research questions namely question 1, 3, 4, 5. So what did we find? Well, next slide, please.

For the first research question, we investigated the factors that influence e-mail addresses likelihood of exposure, and we found that many participants were affected. Specifically 73% of participants appeared in one or more breaches with an average of 5.4 breaches per participant. Therefore, it becomes immediately apparent that most consumers seem to be affected by data breaches.

Using [INAUDIBLE] some regression, we found that the number of breaches associated with an email address increased by 8% per year of use. While 8% might sound like a rather small number, the figure on the right shows how this effect actually builds up over time. Next slide, please. Regarding participants' awareness of data breaches, we found that participants were unaware of the majority, namely 74% of the 792 breaches they saw during the survey, and they were aware of only 80% of them. Next slide, please.

In research question 4, we found that participants like responses show a low concern for the breaches overall as the median was only somewhat concerned. This sentiment was also reflected in the qualitative data we collected as ill

DANIELLE ESTRADA: Thank you, Peter. I wanted to-- this is fascinating and it's really interesting to hear your views and your research on consumers' awareness or lack thereof of their data of the many breaches that at least the consumers in your study were affected by, and then the rules that everyone can play to help them and to mitigate the effects. I wanted to start by asking you about what consumers can do to protect themselves and to mitigate the effects of data breaches, and also what consumers can do in response when they do find themselves to be