w5_2

Tue, Nov 22, 2022 1:14PM 26:57

SUMMARY KEYWORDS

data, consent, people, ai, user, data breach, forgotten, algorithm, week, book, app, australia, means, colleague, company, comment, privacy, optus, ensure, breach

---

00:03

All right, yes. I think with with AI, like it's gonna hang on now. Okay. I think the danger is that like, it's going to take your thaw and life stories.

00:16

So there's always that risk it's gonna get. So I think it really for like, product quality. And yeah as far as size unless it has

00:36

an aspect of it Give, give so that you don't have these days or whichever is other.

00:44

Okay, so how many of you agree with him? Right, okay. All right.

00:57

Oops, sorry.

01:00

Okay. Okay, everyone, you can hear us now. I heard back again. Excellent. So if you just talk to here, apparently, it

01:09

looks like the charger didn't work. So it was completely off. So yeah, basically, that was a very good discussion. So I think our friends here seems to really like the features. But as long as you

good discussion. So I think our friends here seems to really like the features. But as long as you don't share my data, I'm fine with that. Okay.

01:26

The question now is, should we even be marketing people face off? Right or activity place for answers to things that they want? Because like, what they just say, like, oh, based on this person, we, you know, we're going to sell you that item. We're just looking at this, right. It's also like, it could be certain biases that we have based off the person that could be negative. For example, if we realized that this person is a violent person, right, let's promote him nice. And like face everything fine. He's like, Sure, you can have money in that backpack that was off. So it's like, what is the purpose of this AI? And I believe that should be?

02:12

That is a very good, very good comment. Indeed. Yeah.

02:19

So even if I bring my daughter and I've got to want to run on AI, so that a failure is experienced, or something I get as best as I add as a bonus, so quite common for companies support data is they don't really tell us when they're going to delete the data. Data from like 10 years ago, is whenever they're using similar thought on. And even if they don't use it, that just, it just leads to more and more stuff that could potentially be linked in the future. So I like it. The problem is that like, like, if you still want to use the AI, then at least they should have some type of timespan on it. But that's,

02:59

that's, that's a very good comment. In fact, I just met a person yesterday, and another UNSW staff who commented that he was not he was not an Optus customer for the last. He's not obvious customer the last 12 years. But he was contacted as part of a data breach. Apparently, his identity was replete. But everything I've changed, thank God, but the only things that have not changed his his own home address. He hasn't moved home the past 10 years, but the passport and driverless cars have changed. So of course, they can move home, because only because of this data breach, but that was interesting. Your comment that artists should dare to you know, the class is no longer in the database should no longer be database because the you know, this left the company, delete the whole thing. That's a very good comment. Any other thing that you want to share from your discussion just now?

04:06

Is it really a problem? Is it God probably going to be replaced by government agencies? Say like the Border Patrol, something for people visiting?

04:24

Right. So if let's say there's gonna be a change of government that that, you know, wants to see all data from companies and our friend, he says in a situation where telcos need to provide this information, or even a tech company like Amazon, who knows information ethnicity to give up this kind of information. If that has, if that's going to happen, whether that's actually going to still remain a utility for the users. So that actually brings up a comment from our friend here as well. Where do people that company really need to know about our ethnicity to be able to provide better services? There's a there's a really good question. So when personalized information, so personalization, it might in might be right in if Amazon because they have Prime Video. So some, some people might be choosing certain movies because of their background based on their ethnicity. So, or maybe certain products are more valuable. But, but of course, then the appropriate data governance framework needs to be there. And that's where we, you know, my next slide is, I wonder if you can load it for me. Over here. Sorry. So, to get you to work, can

05:51

you send me the slide? I did? Ah, yes. In that case.

05:55

So, um, one thing that was different from GDPR. Ah, has anyone heard of GDPR? Yeah. Great. So stands for the general data protection. General Data Protection, what is the I forgot? If it's an EU, EU law into privacy and data governance. And one of the costs that's very interesting in GDPR. So and this applies to all EU residents and citizens, which means, which means I'm just wondering, any of you here is an EU citizen or resident. Right. So they exist here in Australia. So we cannot assume this do not apply in Australia. They do because we have plenty of EU citizen residents in Australia as well. And GDPR stipulates very clearly that consent has to go came in it has to be very specific. And is the screenshot share at the moment. Yep. Sorry. I'm

07:12

working on

07:15

the projector.

07:17

Block mania, okay. Doo doo doo doo doo doo dah dah dah dah dah dah, dah, dah. Can people see the slides?

07:29

Oh, okay. Okay, sorry.

07:38

How do I see the chat? Where's the chat window? Sorry. Oh.

07:57

That is strange.

08:00

For me, I'm sorry. People online won't be able to see your chat. Huh?

08:06

Why it's just on something to my display because I plugged in the HDMI cable

08:16

that's not it. Don't worry.

08:21

I don't know where you're this night. I can't get to the zoom. Again, it is the rescue

08:43

oh, let's just right here. Oh, don't worry if it gets rid of that. Okay. Go to chat again.

08:58

Okay, sorry, people, sorry, people online. We can't see your chat. All right. If people, if someone can kindly just join the zoom. And then if, if there's any interesting chat, just let us know. Yeah, I can do it. Okay.

09:20

So there's also this content and Privacy Act now. So they're basically the four key terms. The

So there's also this content and Privacy Act now. So they're basically the four key terms. The individual adequately informed before giving content so content to what, what kind of data and what it's going to be used for. And it's found found voluntarily giving consent. And content is current and specific. I can't remember exactly with GDPR terms. I think it's two years. Does anyone remember or No, I think two or three years so every. So once you give a concert invalid for a certain period of time, and then the software we need to ask you to again if you're going to give consent when when that comes And expires. And also in the content also says that you have the capacity to understand what you're consenting about. And what is interesting is the right to be forgotten is also there. Which means there needs to be a capability in the software to say, delete my history, or remove my account, and you need to be completely forgotten from the system. This was actually quite near. So only a couple of years ago that this becomes more prevalent. Yes.

10:36

You verify validate that data has, like a company has probably got because you can't really look at data that doesn't exist and be like, I don't know. But how do you do that? How do you trust them? At all?

10:54

That is a very good question. The thing is, I can't How can we trust companies that they have actually deleted your data. Now in let's say, in in the case of Optus, for example. Let's say they have asked you and your new customer and there's a you have provided this content, and then you have the right to be forgotten. And then you basically say I have to delete my data. They didn't delete it, and now you're a victim of a breach. You could have them use that as a way to litigate Optus, because, hey, I've given you my consent. And I have asked you to forgive me. But my data is like only in that kind of cases, then you learn? Unfortunately, I don't know, to be honest, at the moment. What is even more complicated is machine learning models are trained on a lot of aggregate user behavior data. You're one among the many data points. You and I, and this historical data are used to build a personalized model of some sorts, right? How do we get a machine learning model? To forget you? That is still a difficult question is not yet resolved is still a research topic. So it's a research topic. So it's not even yet resolved. A model that has been trained on many years and one person to leave the system say forget me, how does a model getting untrained on that person alone, for example?

12:34

Still very difficult.

12:39

Okay, we don't have much time, I will not discuss this. So one thing that I need to mention about Cambridge analytical, I think we all know about this Cambridge analytical, let me just put this here are the Cambridge University researcher use a Facebook app to basically collect data from people taking a survey on Facebook, and accessing not just their own data, but their friends. So so the data is harvested from millions of users all the consent are only provided

from specific users, not all of them. So for example, SAP is a friend in Facebook, and Seth has given consent to this app. But that that means concepts also providing consent for this app to my information about me because I'm connected to serve. And then these actually data will use for targeted advertising, even for political campaigns. This was this became a big issue. And the main issue was the consent process itself, it was incorrectly wasn't specific, it wasn't transparent, it implicated not just me, but also people connected to me. So that was not supposed to happen. Your consent should only be given by you, the user, you don't have any authority to provide consent for people connected to you. In Facebook, and, and therefore, in terms of let's say, Australian Privacy Act, there are things like, you know, you need to review your data collection usage policies, and you need to have struck this balance between the privacy protection and innovation. And one thing that's important is not self sufficient to merely follow the letter of law which means again, going back to my earlier lecture, what is lawful or legal may not be ethical? What what is your stand as a as a computer scientist, IT professional Do you want to to be someone who's ethical it not just like, Okay, it's allowed by the law. Back then it was allowed by the law. But in fact, in here, the law was playing catch up after Cambridge analytical process. And they realized, hey, this constant process is not good enough. Because it can be shown, like you say, Oh, we've got consent we can get, you can do this kind of analysis. But it was not ethical. Okay, I'll skip that there's not one controversial study. Let me just talk about this briefly. This was a conversation article, I wrote with a colleague in the NSW Soliel, a couple years ago. The day was 2018, I think

15:42

2018 Can't remember.

15:46

So, theory tracking, I want us who actually have used it in the past, for the P female in here, I have used them in the past. And I realized, this data can be used to know can be this data has been sold, and then mined by advertisers as well, including Facebook. So if, if, if there's a menstrual cycle, and you know, women in their menstrual cycle, closer the menstrual cycle, the more moody they could be, and it could maybe lead to certain things that they might like certain things more. And this, Facebook can then have a targeted app based on your menstrual cycle. So there's many apps of like this, and one of them is called flow. What was even more surprising. I just realized, but also a couple of months ago, when, when this abortion law was passed in United States, several states in us. Apparently, this app data back going back to your statement about government wanting this kind of data. This data could be used by government, in those states, to find out if someone had an abortion and could be used against you. So there needs to be a proper governance about this kind of data. And therefore, you as a user, you might want to consider whether I should you use this kind of app, or not? The best thing, what I always advise people is if you want to develop a personalized system to make features, you know, better features for the user, make sure the data stays in the device. And even if you need to build models, the models, you know, the models have to be trained on data staying on the edge. And, and there needs to be a way to propagate to allow an anonymity in a way that in say, for example, algorithm like K anonymity, if there's a unique individual in the crowd, do not use that unique individual data, because then it will be very easy to trace the data back to the individual. You only want to use data from the crowd that has large group of memberships. Otherwise, how do you ensure privacy of the user? There are many, many things

we can talk about this. So I'm more of a researcher in the personalization and machine learning with time series and sensor data. So my colleague, Sally is the one that our colleague is the one who is more an expert of privacy, and when we wrote about this together, and that's why now in companies, we also need to ensure the data safety, security and governance so so what is the data security governance, so it's a policy around data for operational purpose in our organization. And basically, if it defines actions, whose results who is responsible for what action so a good governance framework will ensure three things data quality, their quality means when data arrives to you, you want to make sure it's pre processed properly. So you may want you may not want to keep the raw data. If you're, if your company has a proper data governance framework, you don't want to keep the raw data unless it's only useful identifying purposes and they may need to be stored in a special database. We're very strict authorization access anything else? If you're gonna use it for modeling, lift the identity out? Just use everything else. So the day quality has to be there. Data Availability, and data protection, how is the data protected if you don't have a data security governance framework in place, and if you don't follow it carefully, there's a issue of things like data breach. What is interesting is you can look up IBM, IBM does a yearly report on data breach in Australia yearly report. And I'm sorry to tell you that the figures goes up every single year. I just checked it last night, it has gone up to about a couple millions from 2021 to 22. And so and they haven't included this report was two years before the Optus breach. So I'm not sure what the figure will look like next year. And of course, it's a massive cause of the, you know, from the loss of reputation of the company. So there was a mandatory data breach, for example, Equifax data breach.

21:00

All right, so I think I have to stop there. I just, I'm just going to quickly Alright, let me just quickly go through this one first, before going before stopping. So, one thing is one of the one of the major issue about this algorithm is because our algorithms are becoming less and less transparent, especially those who are actually AI based, and one of the highest number of keywords in in these AI frameworks, ethical frameworks, blah, blah, blah, is transparency to ensure trustworthiness. Transparency is a key requirement. It's a it's a single most common principle in a global level. So, this is a study from Latin enhance. So, then this goes to the question whether we are deploying a black box or white box algorithm the term of black box was white box was actually first discussed by the father of cybernetics, Norbert Wiener 1948. On this book, control communication in an item on the machine, actually not a book, so not article is a book, you can access it online, if you if you want to read it. It's quite interesting. Book from 1948. It already introduced, what is the difference between the two. But in fact, basically, if it's a white box, you'll be able to inspect input process output, you'll be able to understand why the output came to be from the input. When it wears a black box, you just have to accept that it's actually something magical going on within the box. And it gives you a certain output. So there's a stronger need for transparency. Excellent. Now Explainable AI. Again, why because there there are a couple of different articles are internationally. I just pick up one for example here, customers rights for explanation GDPR article 15 requirement for him in a loop GDPR article 22 requirement for algorithmic auditing us algorithmic Accountability Act. Australia is not yet listed. They're still playing catch up. But it means we need to make sure even we have deployed machine learning algorithm in our system. They need to be explainable. Mike, that's my colleague, Jacqueline Fein from University of Melbourne law school. So he is one of the investigators in our center. We will talk more about transplants expert in a week 10. So we have a Microsoft Caspar SOCO will give a talk about this, he's an expert on this topic. And next week, I'll talk about fairness. So, just very quickly, why fantasy is important. This was an a book from Sofia novel algorithm of oppression. So for example, a search of

women back then, there was a largely white women or even girl and this was three years ago search engine result on SEO what you could see is male, white, right. So people wearing ties. So this is basically the bias of such and such changing. So very homogeneous. There is no diversity at all. It was those are the things that were discussed in Sofia nobles book, algorithm of oppression is just perfectly the biases in our systems today, so this was last night I did a screen screen capture I think they've done a better work now it's more diverse people of color more female not just male. And I will also talk about this briefly in the I will also talk about this I'll actually Virginia Eubanks is coming to Melbourne next week and the week after I'm going to attend her workshop that will be very interesting. She basically talks about this a lot of tech for good design for social services. But instead of actually achieving the good is actually creating a further inequality um, yeah, I think that's basically it.

25:46

Girls don't share their interest let's even get back to this zoom chat. Right. Okay. It certainly wasn't

26:00

going to be good. Yeah, so that's where the readings if you're interested to learn more about FASD okay Ah, sorry reminder next week so next week maybe you don't have to be reminded but yeah, just in case not even lecture.

26:33

Oh can you give me one second? Yes, yes. To the shut the meeting down. So the reason I want to shut it down now is it takes a while to compile into a video on my hard drive, and then I can start packing up last. Alright everyone online. Ciao.