w7_1

Sun, Oct 30, 2022 10:08AM   1:01:31

SUMMARY KEYWORDS

bias, people, fairness, system, decision, data, world, based, ai, anchoring bias, problem, technology, book, software, images, cognitive biases, transparency, google, compass, arrested

---

00:13

Send your star

00:21

thank you so much I think but now I

00:33

really need an extra month because I need to charge Sure so I'm gonna pick up over here I'll be on the on the corner

00:55

easy to fix easy to get. So, people fixing problems here this is Business Solutions Engineering right here right now. So

01:12

worse us UX is actually is actually Hi everyone

01:38

all right do you want me to run out and get

01:48

36% Okay thank you I thought I was doing so can I invite you into your

02:15

and if you try email you the Zoom link

02:18

is okay and do

02:52

let me know when you have joined?

04:29

Somebody say again

04:33

ah no because I mute myself from here

04:42

all right can you hear me hello people behind can you make your no response?

04:51

No, no, I'll take mine.

04:55

share my screen. It's up to you Excellent. All right people alongside us yet Great.

05:46

Thank you so much, please.

05:47

Oh my god. You brilliant. Thank you.

05:58

Today, we're gonna talk about classical relatively big.

06:03

Okay, got it. All right, I'm so sorry for the delay. No, I don't want software update way.

06:14

So we just finished staff meeting today actually haven't finished. So we left early. And, you know, we had sought out a powerpop issue. So thank you everyone for being here.

06:29

So let's start.

06:32

I'd like to acknowledge the medical people, their traditional custodians of the land, and also, like, pay my respects to the elders both past and present, and extend that respect to other Aboriginal and Torres Strait Islanders who are present here today. Okay, so I think I'm just quickly so today, I'll just do a very quick recap of where we were last week. I think we are stopping at the aspect of stopping just a quick glance on fairness and transparency. And then today, I'll, we'll have a deep dive bias and fairness, so this is one of our research topic in our group these days. So just again, I like to recap on transparency, then I think this is also a very important issue to discuss before talking about bias and fairness.

07:37

So um,

07:39

so I mentioned briefly last week, before we run out of time, that transparency is really a key requirement for password EI. And, and it is, it is actually highly demanded in, in EU court of law and also a lot of legislations and AI code of ethics that have been released all over the world. Transparency is one of the important things to look at. But of course, as we all know, the way we design our automated decision making system a BMS or AI based systems it can be designed in either an a black box, metal, black, white box, and these actually, these terms of black box and white box are not new. So this was first discussed by Norbert Wiener in 1948. So in his book cybernetics, so he actually mentioned how we can actually create a very clear system of software that can do that, that the not just the designer, but the user will have a clearer clue on how things work, and how can you use it to control different parts of the world so so in, in other words, if it's a white box, you'll be able to see it's more transparent to be able to see the inner workings, you will be able to know the the decision rules that underlies the output of the system. Whereas a black box, you know, you just have to accept there's some magic going on within the box such as for example, these days using deep learning and then we we know that this is like the most accurate, the most accurate output and we just have to accept the outcome. Now, the problem with that of course,

09:49

with the

09:51

with system becoming more opaque the requirement for transparency It is becoming harder to come for now, what in your assignment asked you right to look at some, some hypothetical or real system technology that you'd like to deploy in a specific context. Now, actually, this, this is actually a very real world setting, because when you are working after you finish your degree, you work as a consulting, you're going to cost a consulting company, you may have to decide which software or which AI bases and which feature you want to you want to use, then you may have to develop a system of systems to automate decision making in your company. Now, the thing is a lot of these ml systems they because of trade secrets, because how they definitely is not that clear how they work. And, and in fact, to make this kind of system to be equitable, and fair and not bias is even harder. So I think this is why I want to touch on this before we go further into the fairness and bias. And is important, of course, with different GDP articles that we discussed last week, for example, customers, right for explanations, it is important for us to, to look at the issue of transparency. And we will have one lecture dedicated just on interpretability and explanation, AI explanation, and there'll be on 16th 16th of November. So my postdoc, Casper Sacco, who's actually an expert in this topic, he's got tons of papers on explainability. So he even have explained explainability fact sheet paper that's actually now already highly cited. He'll give a talk about that his whole lecture will be on explainability. Now, so let's jump into payments.

12:07

Sorry, this is not COVID it's just

12:12

it's just the season of the year. Spring. I've got so much different allergies to different types of pollens. Now. And in Sydney. I'm a lot better in Sydney. So in Melbourne is really bad. No, because I actually no, I've just moved from RMIT to unis. W so yeah. So whenever springs in season, I struggled quite a bit. Now fairness. The topic of fairness has a lot of different facets to it.

12:50

Sofia novel.

12:53

There's so many books. So one of these, one of the popular books on this topic are words of oppression, written by Sofia novel, talk about how search engines reinforce random racism. And this was basically looking at aspects of like, say Google, Google search. If you're doing you know, you start doing a search, and it does a a suggestion on basically a query completion. And, for example, the one that you see in the snippet here. Why are black women and such suggestions by Google back then when she wrote this book is why are black women so angry? Or why are black women so loud or so mean? So these are the suggestions coming out of the the the text of corpus that Google was training on? So so if you if you see

14:00

these pictures,

14:02

this was taken a couple years ago, on a picture of Google image search on women, and girl and you could see predominantly white. And also, three years ago, when I started doing a lecture on bias and fairness, this was a snippet back then on SEO. And take a look at the difference between three years ago. And now. So this was taken last week. What are the clear differences? How has Google actually rectified it? So since the book of Sofia novel when it was released, and there are so many other books that were written, including Kate Crawford, books, articles, companies, like Google, for example, have become more intentional in correcting the biases in their own system. And so what are the key differences that you saw in the snippet here and the previous picture?

15:11

SEO,

15:12

Search Results image results three years ago, and now? Have they managed to change something in the representation?

15:22

What do you see? It's different. They look more diverse.

15:33

They're looking more, you know, at least now they're more women. So, previously,

15:38

more people of color in the chat

15:40

are people of color. That's right. That's very good. So, um,

15:47

so that's just one example. But I like to start with the aspect of bias. So where do they actually come from? So why do we Why do this bias come from when we think, if we have already come a far from the first lecture, when I mention technology is not neutral. And they could be interest bias in USANA. So where do they come from? If we think about, if we actually think about as software system as input processing output, right. So either it could be, you know, so you normally have a code, it takes an either maybe just one variable as an input, or an actual input data database, or even a whole corpus of images or tax. And then you might come up with a process, which you know, basically your algorithm that you write, or it could be an ML algorithm to actually learn how this entire process works, and then retrain it based on this to be able to mimic the same system. Now, the bias can actually come in each of this part of this system, either in an input or in a process when you're actually coding it. And also, even when, even if the, if the system itself is not bias in the way it's conceived. But when the human users are interacting with the system, there could be bias from the user itself when interacting with the system. So sometimes the technology, we can always blame the technology alone, but but the bias comes from the human computer interaction with so when this interaction happened. So let me just

17:35

give you an example

17:37

of why this bias can come from different different parts. So has anyone heard of this case? Study? beauty.ai? No, so this was the first international beauty contest judged by an AI. I think this was done in 2016, or 2017. So what happened was this initiative by Russia and Hong Kong base lab, it's actually widely supported. It didn't get game sponsorship and support from Microsoft from Nvidia. So hundreds of 1000s of people submitted photos, their own photos, their their selfies, basically. So they say, you know, submit your selfie, and we'll get AI to judge how pretty you are. Back then, it sounds pretty cool. So people did,

18:35

and what was the result? Um, so it was

18:42

actually very interesting that the people of color were hugely discriminated, hugely discriminated. So. So they have algorithms that are used to judge you know, people based on their symmetry. So I don't know if you can see the features here. But the features here include things like you know, how Symmetrix your face on left and right, for example, or looking at for example, where do you have wrinkles? What are you know, how you look at your age, what age actually made you match your true age, and, and things like that, the skin color was not evaluated, but for somehow out of the winners, you know, the top 44 who are considered the most attractive only, I mean, only about six were Asian, and they were actually none who actually have dark skin. So, what was the problem there?

19:51

So, the problem back then was I think,

19:56

people only realize when they look at it, they The submissions the HA submission from Asian countries only constitute 5%. And submissions from the African countries are less than 10%. Less than 1%. And, and because the, the algorithm, I'm not used to seeing pictures of people actually having data columns, different skin colors, it did not actually see them as attractive. Because if if you know how these kind of machines are trying, anything that doesn't have equal representation could be treated as noise. So in fact, that's why what happened. And also, it adds a complexity of because it's based on also annotation. So human input, who manually annotate attractiveness, it wasn't really discussed, it wasn't really clear whether or not an audition process was the beauty.ai team did not reveal how that actually was done. But a lot of people suspected, it may well be because there's so much submissions from the US and European countries, the manual annotation that look at attractiveness, mainly based on how these people tend to be attractive, and potentially the annotators may only just be white. So this is really a problem. Right? So if you look at this beauty.ai, where, where do they come from? If we think about an input processing output of a software?

21:55

Where do this bias come from? From the input data?

22:02

And not just the input data, but how are they annotated. And we always know that these days, I think after I think we've come so far, in understanding, diversity inclusion, especially in this country, that we know diversity of opinions always lead lead to better results. But back then training, these kind of models. diversity of opinion will never be was never a thought. So they thought, okay, we can just recruit a group of amateurs. And maybe they could be a bunch of university students from the US or from wherever, and we'll just get them to think which one's pretty, and even if usually recruit certain, you know, certain this was done by companies. Right, Russian, us. And also, maybe they are and also Hong Kong. So let's say we, you know, some of these are from Hong Kong, because because of the way how the whites are portrayed in the media, they actually might have a better idea of what beauty looks like potentially in their in the eyes of beholder, as portrayed in the media, but not so much of people of color. So we know that bias come from the input into this kind of system. But interestingly

23:38

it may also be the other way around, even if the software is completely fair and its representation. When it's deployed to the real world, we human beings, we have our own biases. So the way we use the software, actually interest but bias for the bias in the software itself.

24:13

Let's see this one here. I don't know if you were part of this experiment several years ago, Quick Draw experiment. Did anyone try this? When it was released? Yeah, okay. So remember that. So what happened? What did you do?

24:33

What did you do during the game?

24:39

Or something and then they credit recognized what it is. So you're something so Google released this game, you do have something and and then the machine, try to recognize what it is, as someone else raise their hand.

24:59

No on tribal withdrawal experiment, remember what you did is draw an image of maybe yes. Yes.

25:11

Pam also did. And how's the recognition back then? Was it accurate?

25:19

Pretty close. Cam. What did you draw? Cam is online? Do you have a speaker me I mean microphone on you the audio might

25:34

not come through.

25:36

I don't remember what I drew. Okay, let's see if we can play this YouTube video

25:49

or if I can share audio from here.

25:54

No need to be logged in through your MIDI jack into the

26:02

Okay, no problem. I will play the video.

26:08

Okay, but if it is,

26:11

I will have to stop sharing and I think I have to share my screen with the

26:17

audio option. I think that's what it worked, how it works. Let's see. Share sound came

26:40

close your eyes and picture a shoe and people okay, did anyone Picture This? This? How about close your eyes and picture a shoe and people okay, did anyone Picture This? This? How about this? We may not even know why. But each of us is bias toward one shoe over the others. Close your eyes and picture a shoe. Okay, did anyone Picture This? This sorry? How about this? We may not even know why.

27:08

I

27:10

was alright, this might be okay. Okay, now you see.

27:22

Close your eyes and picture a shoe. Okay, did anyone Picture This? This? How about this. We may not even know why. But each of us is bias toward one shoe over the others. Now imagine that you're trying to teach a computer to recognize a shoe, you may end up exposing it to your own bias. That's how bias happens in machine learning. But first, what is machine learning? Well, it's used in a lot of technology we use today. Machine learning helps us get from place to place gives us suggestions. Translate stuff, even understands what you say to it. How does it work? With traditional programming, people hand code the solution to a problem step by step. with machine learning, computers learn the solution by finding patterns and data. So it's easy to think there's no human bias and that just because something is based on data doesn't automatically make it neutral. Even with good intentions, it's impossible to separate ourselves from our own human biases. So our human biases become part of the technology we create in many different ways. There's interaction bias, like this recent game, where people were asked to dress shoes for the computer. Most people drew ones like this. So as more people interacted with the game, the computer didn't even recognize these latent bias. For example, if you were training a computer on what a physicist looks like, and you're using pictures of past physicists, your algorithm will end up with a latent bias skewing towards men, and selection bias. So you're training a model to recognize faces. Whether you grab images from the internet or your own photo library, are you making sure to select photos that represent everyone? Since some of our most advanced products use machine learning? We've been working to prevent that technology from perpetuating negative human bias from tackling offensive or clearly misleading information from appearing at the top of your search results page.

29:27

All right, I think that's enough. That's a question. Yes. When was this video that this was made a couple of years ago? Before all the time when?

29:40

So this video we made three months before Google admitted that they accidentally made a racist algorithm that selected black people's faces and said they were gorillas.

29:51

Pretty much before.

29:53

So yeah, so that was actually highlighted in the book by Sofia novel Thank you for reminding me about that. Tom. I don't know. Did I mention this last last week? Two weeks ago in the lecture? No, maybe it happened. Right? Sometimes I forgot to mention something. So what Tom mentioned was in the book of Serbia novel, when, when people are searching for Gorila it in Google image search, it also picks up people of Blackhall up as shown as a result. And because they couldn't actually mitigate that issue, they come, they completely remove the term gorilla. So if somebody search for gorilla during that period, net can't find anything, because that's such a completely remove. So it's not no longer something that could be recognized as a as a term. Until they tried to rectify it. So yeah, so they signed it, they tried to mitigate this. But anyway, now harbor stops sharing this and go back to my

30:56

slides. Okay.

31:04

was giving me some reason your slides are not coming over onto my computer. So I am trying again,

31:09

here they go, because I unshare that, and I just share this.

31:16

All right. Now, so if you look at what was mentioned, there was basically,

31:32

you know, because we all have our own biases, the way we draw something also could mean, you know, the different representation of less than what a whale is, or a shoe is different to different people. So there's culturally, we have our own specific contextual biases. And and this is where I think, in terms of we interacting with the system, is where cognitive biases come in. So there's hundreds of terms of cognitive biases that have been coined by a psychology experts who are looking at biases of humans. And I mean, this is a snippet of cognitive biases from the decision Lab website. I really like this website, because it's very interactive. So if you click on one bias, you actually give you a couple of different examples. So a summary and example and how these biases relate to another bias. So feel free to take a look at that. And in fact, decision lab bias is here is only a subset of a longer list of biases in Wikipedia, because the one in decision lab, they focusing a lot more on how biases can influence behavior economy. So for example, if you want to design a system that looks at the economics of you know, what, what is the incentive for stakeholder a, in order to do to purchase a product? And how do you actually incentivize it against stakeholder B, so you can look at the different biases and then play against them play for one or against one another. So this is basically a how decision lab has actually come up with their own set of biases, even with their own set of biases, there's still more than 100. So one is one of the most popular bias biases that you and I have. Have actually, I mean, experience, maybe we don't realize them is called anchoring bias. So this is one of the most pervasive cognitive bias. So anchoring bias is a situation when you're receiving information. And the first time you receive it, you accept that as the truth that you will anchor on. And because you receive the information early, that becomes a point of reference. So any other information that comes afterwards, you try to skew it towards that first point of reference. So again, from decision lab, I realized this image so if you saw a mock with just a price tag of $300 for the first time, you know, it may not do much for you. So you you know I don't that's very expensive. But let's say the first time you see the mark with the price, DEC $1,000, right. You say Whoa, that's a lot expensive $1,000 for a month, but then the second time you saw it again And you went to the same shop. And it's, you know, a big discount, it became $300 and say, Wow, that's a, that's free money. That's, that's a great discount. So with your anchoring bias, because you had a point of reference, it was $1,000. Last week, and now $300 Time to buy, I'm gonna buy it. So you don't know, maybe it's actually the half plan, or it could be priced as $300. But, you know, they tried to put it up first before they wrote it down. So this is anchoring bias. And anchoring bias is so perfect isn't even the way you read news, the way you engage with social media, you know, as I say, you might think I designed the software are designed is to be completely neutral. But because the way we have, you know, we have different biases, including, for example, anchoring bias, we stick to this information. And let's say, you know, you're doing research or your essay, which was already, which I have already submitted now. And you might have read one literature that you think this is really good, I'm going to anchor everything based on this. And there's a danger in that, that, you know, you can only you cannot really accept everything else as a source of truth.

36:26

And this would lead to things like confirmation bias. So confirmation bias is like you seeking to, to get a confirmation from what you believe is true. And this is where things like misinformation happening, you know, people got into the trap into the loop of fake news, reading fake news, let's say, you know, if you get injected with Pfizer vaccine, or any any of those mRNA Bill Gates will have a chip inserted in your nose. So for example, right, and, and there was already say maybe a scientific article that shows you this, this is all in one, it may not be peer reviewed. Remember my first lecture, I mentioned, there were so many debunked research that actually was completely false. Like for example, how 5g can cause you to have COVID And that was actually submitted to a biomedical journal. But you know, thank God that was actually rejected and but the archive is not never pulled out. And people use it as a source of truth. So people try to seek confirmation based on what they first anchor on. So a lot of these biases actually relate to each other. So another popular one is called implicit bias. So this is taken from Google the example I'll show you later but basically implicit biases. We all have our own mental models and we make an assumption based on our own experience, how we are brought up in Our.

39:31

Hi everyone, we're trying to find floor. Hey, hang on. Hi everyone online the university's IT department have decided to hack into Florida's computer and and force her to do a zoom update right now in the middle of our lecture. If everyone would like to send an email about this complaining to the university I think that would be pretty fantastic clearly I'm going to do the same thing I'm terribly sorry it will be okay flop for

40:42

you make me heart okay. So how far was I gonna like people online tell me Okay great. Okay,

41:19

share screen again. So class share screen came in ecommerce, please

41:25

give me one second.

41:30

Sorry,

41:31

I thought you'd reintroduce.

41:36

Some people online, will you there was either when actually mentioned about implicit bias or don't have to repeat it.

41:48

You should be able to shift

41:53

Did you manage to hear about implicit bias?

42:00

Anyway, right.

42:04

Yeah, so I talked about implicit bias just now. So. So, remember, your mental model is different. So how would you? How would you mitigate against this? If you're designing a technology? And you want to be able to account for example, implicit bias?

42:27

How would you do this? Yes, research is more recertified by whatever, computer, you know, processing.

42:47

So our friend here says research research, research on the user

42:52

research on the use in like the domains of for example, for your gesture, maybe instead of just looking at the gestures that you know, research, like gestures all across the world get as wide of a data set as you can.

43:04

Yep. So if it's about gestures, then do more research about gestures. That's great. And I think user research is very important in this case. So yeah, there are a couple of ways. So importantly, always remember before you deploy anything, user is important. And that's why I actually started even this. You know, in my first lecture about user research, it is very important if you want to ensure we don't actually induce bias in our technology

43:45

now attacks case study compass. So compass is a very popular was a very popular tool that was used in courts American course. I'm sure many of you have heard about compass.

44:09

Okay, so yep, not not surprising. It's been described, discussed deeply, widely in a lot of books. So compass was there who by no point in 1990s, so it wasn't AI based, it was actually statistical based. So I'll categorize it under ADM, automated decision making system but it doesn't use AI is designed to assess the risk of recidivism. So what is recidivism? How likely a defendant will commit another crime after being released. And therefore this is normally used by by the judge to make a decision whether how long the sentence should be, and whether this person can be released on parole. Now, let's look at some of the outputs of this contract.

45:01

So it was already widely deployed so we've got Vernon and brucia on a picture.

45:17

So, Vernon was rated low risk and British us rated high risk. Now let's look at the actual data behind Fernand Berisha. So both are actually arrested because of tap, but burnin is prior offenses include two armed robberies, and one attempted armed robbery. And then up and during court, it was you know, it was categorized as low risk, and therefore can be released after you know, use a shorter offense, shorter imprisonment and then can be released early for parole. And after that they actually for now actually had a subsequent offense, which is a grand theft. Now, what about Bucha? So Bucha actually took a kid's bike and scooter that was sitting outside because she thought, oh, no, nobody needed him. So I just picked them. So those are very minor tariffs. But she was categorized as high risk and had longer sentencing period in comparison to

46:40

now this was the the AI. So the judge that potentially use this completely being misled by the decision. Another example,

46:56

I'd like you to actually look at the attributes why what makes this could potentially be

47:01

a bias, grab possession, arrest, we've got done, we got bettered again, Dylan on the left, categorized as low risk. And Bernard on the right, considered high risk.

47:21

So Dylan forgit, rated low risk after being arrested with cocaine and marijuana. And he was

So Dylan forgit, rated low risk after being arrested with cocaine and marijuana. And he was actually arrested three times on drug charges after that. Now, what happened to burn what was actually the actual

47:42

offense that he took the data shows Bernard actually was just resisting arrest without any violence.

47:57

It was He was considered high risks, and he had no subsequent offense offenses.

48:03

Now, out of those two examples, I'll show you what are the possible possible

48:12

features that led to the bias in this ADM tool just to give you a bit more information, in addition to some historical data, which of course should be pretty straightforward, which is just the data, the compass data is a compass is also built on heavily on questionnaires. So, so the offenders will ask to actually feel the survey and their questions, you know, I will say very sensitive, including things like

48:59

have your parents separated? And how old were you when your parents separated? There are also questions like how many friends how many of your friends or families have been arrested before? Now, these, these are the ones that actually

49:28

the the researchers thought could all could actually heavily weighed on the decision making tool. As you can see, based on all the what's shown here,

49:39

those actually having

49:42

a lower or lesser degree of risks to the society seem to be shown as high risk because of their a lower or lesser degree of risks to the society seem to be shown as high risk because of their color the skin

49:56

and it was very unfortunate

50:00

that it that the system is proprietary. So it were our I had to refer back to my first statement about blackbox systems. And a lot of these company built software, they are not open source. And it is very hard to actually know the, the, the inner workings of the algorithms unless you test it. So you could do it by testing what the decision outputs look like. And you compare the results. So this is what the ProPublica people did. So in those articles, if you can actually read it, it's a really deep dive into this problem complex, that they actually have to look at a lot of the data and they realize that the survey data became, seem to be the proxy variables that determine the risks. And also because it's actually hardly related to the photo. And the photo of the skin color became a really strong determinant of risk. And that was absolutely discriminative. So that's an example of technology that has been deployed has completely gone wrong. So compass is no longer exists now used to be but not anymore.

51:33

So what is the difference then between bias versus fairness? People actually think, oh,

51:41

you know, do they mean the same thing? Are they the same? Are they different? They have a lot of overlap. But if you think about the word bias, that should focus more on our representation, either representation in the software itself, in the system, in the data, or the bias within our own mind, how we treat others how we treat the system. So bias has to do with the representation, how we view the world how to solve our view the world.

52:17

Whereas fairness focuses more on the decision outcome. So the outcome division outcome means do we treat everyone equal?

52:31

Or, for example, if let's say it's a decision, output of a system is a binary, let's say binary in a way that what if you classifying Sebastian with a credit history, and Florida with my own credit history, and Commonwealth Bank deploy a system that looks at our loan mortgage application?

Right? And should we should we actually approve Sebastien loan application for a new mortgage? Or not? And what about flora?

53:16

We'll look at our history, you look at all of the data, the dishes in

53:21

we can say that this mortgage, this classification system, we only be fair, if let's say wondering criteria, because there's so much different metrics of fairness. And I'll talk more about it after this after the break. But for example, if we say equal opportunity, that means Sebastian and I must have the same opportunity to get yes or no, it has to be 5050 This system could be completely unfair, if the it favors against the women for example, or against the elderly or certain age group. So it looks at the decision outcome.

54:10

So so if you look at the famous

54:16

one of the first paper on this topic, which I mean, I really encourage you to read if you want to learn more, is this paper by think on a super name. Very last name. Suresh. He's, he's a world leading thought leader on this aspect of computing fairness. So there's two different worldviews What You See Is What You Get well of you and we are equal worldview.

54:49

So, so the first worldview is, well, the world is somewhat not fair. We'd have to accept that.

55:03

So all you have to make sure is we capture the data to represent the world as it is.

55:11

And make sure we achieve what you see is what you get worldview. So let's make the system

55:21

that the way when we sample our participants, the way we sample our data, the way we sample that, you know, let's say if it's good with Compass, then make sure those with darker skin color and a lot of skin color there, you know, they will present a world if it's if the world is predominantly white, we have to make the system to use predominantly white images, for example. So that is why this is when you get we are all equal worldviews. Looking at

55:55

Do you have a question?

56:00

It's not about selectivity. Okay.

56:05

No problem. So what we are all equal worldview is, is trying to separate the issue of well, you know, what, how the data is representing the world. But looking at, depending on again, the fantas match features, which I will talk more about after the break, is we need to ensure there's an equal equal treatment, equal opportunity, or there's no disparity among the members of the world. And how do we even view the members of the world that's a game and our discussion of a break, whether we view them as groups or we view them as

56:44

individuals so so based on that,

56:57

let's look at this paper called Gender shapes, again, very popular paper. So they even have their own website, you can look it up on Genesis dot gender sheets.org. So so this is a very famous paper, by joy, and Timnath Uber,

57:19

back then work in Google.

57:23

And basically, a back then facial recognition, technology was pervasive. So every company has their own version of it. So Microsoft has their own IBM they own and they basically look at whether these because the author's they actually have darker skin color. And he, the first author actually tested this first on herself. And then she realized that the, this facial recognition system didn't quite work on her. And that's why then he she actually picked this as her PhD topic. And this was actually one of the papers out of her PhD, then explored it and she actually tested this on a lot of different images. And whatever look at is basically software already deployed in the world, even from IBM, Microsoft, they perform best on lighter screen color faces, but it performed really poorly on darker skin faces. So, if you think about this, do you think this is more of a bias problem or fairness problem?

58:49

Yes, bias bias. If you think

58:55

bias right hand if you think fairness left hand if you think both both hands

59:06

okay. So,

59:10

someone who say fairness, can you say why this is apparent problem

59:24

is what we are concerned with over there is like, the AI algorithm doesn't detect faces of people as well. The outcome is what we are concerned with, in fairness, fairness problems. You think about a possible outcome. If this doesn't work. You think of a context where these could actually affect them negatively.

1:00:07

Where will facial recognition technology use

1:00:14

this sample and you want to enter some for other countries, you have to cross the border and some are currently offering like your scan your passport, and they do a face recognition, whether it's correct.

1:00:31

Very good, thank you. And the airports, we all have used them. If you have a travel recently, everything's automated now, even in the Australian Border.

1:00:43

So if this don't work, then you have to use full off people have darker skin colors, which will really bad.

1:00:57

So that if that happens, then that's really to do with

1:01:02

fairness. So until it's

1:01:07

a time when it's only looking at classifying, you know, classifying A versus B or whatever. At that point is very much to the issue of bias. But when it's actually start to be used to make decisions, then the issue comes into fairness. So that's why this is both a bias and fairness

1:01:28

problem.