w8_1

Tue, Nov 22, 2022 12:52PM 53:30

SUMMARY KEYWORDS

people, harms, technologies, autonomy, principle, crime, autonomous cars, computers, humans, questions, commit, ethical frameworks, fairness, facial recognition, reoffend, judges, decisions, duplex, probability, machines

00:17

No no you don't need to know that way Barry you're gonna fix that in just a moment now if you could join that Zoom meeting that would be fantastic I haven't just emailed it to

00:37

you you you join the Zoom meeting and then I'm going to make you co host you see

01:07

now I'm making a co host. Okay

04:11

people on Zoom Would you mind typing in if you can hear me over the microphone now.

04:17

Okay very good. No they can hear the people on Zoom. Can they see us can you share your if you go to see ah chefs we don't have to zoom has to be given

04:59

permission Sorry I have to quit zoom to have the permission from my screen soon will not

05:05

be able to call until it is quit. Okay, that's fine. Yeah. Leave me think. Mute Okay,

06:58

there we go much better.

06:59

Yes, yes. Sorry about that, that's

07:13

okay, if you do want to? If you've got a burning question, please do stop me, I'll be very happy to take questions, there'll be chance, the end of the first downbeat of the end as well. So, I've spent the last 40 something years, I've tried to think about it now, working in arts pathologists, and for most of that time, really, very few people that cared about what I do. And that was, okay, that was reasonable. That was, that was not a problem, because a lot of what we were doing was very abstract. It wasn't impacting upon people's lives. But in the last couple of years, people have started to take notice of what's going on in artificial intelligence. And it has started to impact upon our lives. And it's the point now where you can't open a newspaper. I haven't opened the Sydney Morning Herald today, but I'm pretty sure that there won't be just one story that matches US intelligence. There'll be multiple stories about how artificial intelligence is touching our lives in some way. And sometimes in good ways, and sometimes in bad ways. And so increasingly, I get rung up by the media to try and explain what the technology is doing, where it might be taking us. And I want to share a moment. One of those telephone calls, I think, was in 2018, where I had a real moment of revelation about about this about how I was starting to encroach upon our lives. So it was

09:01

it was a story that broke

09:04

as a consequence of Google's i o conference every year Google has a big conference in in California, where they demo the latest, exciting stuff. And there was a very interesting demo that happened in 2018 for Google's newest chatbots called duplex. If you haven't seen the demo, I strongly encourage you go to YouTube and type in Google's duplex demo IO. And it will give you the video of that it's an in very impressive demo even today and back then it certainly caused a lot of reports that chat bots chatbot has been trained so that you can do various that you can be an assistant personal assistant for you so you can ring up a restaurant and book a table. You can ring up a hairdresser, and book your appointment. You can do simple things like this. And the the conversation, isn't it? We'll have a proper conversation. So you ring up the restaurant and they'll say, Can I book a table for four people at seven o'clock? And it says, we don't have the if the person says we don't have tables at seven o'clock? How about 730? We'll understand that and say, okay, 730, and then it will ask, you know, can we do cater for vegetarians, and it'll be able to understand, follow the conversation along like that. Now, when the demo happened, people were pretty impressed. It's pretty capable, making these conversations. But equally, people were pretty distressed. Because the duplex was easily mistaken for a human. And indeed, duplex undenied like a human. And so not surprisingly, lots of journalists rang me up, and we're worried about computers pretending to be humans is there's actually plentiful Hollywood movies where computers are pretending to be humans. And it all goes badly wrong. So anyway, I was answering questions about what chat bots could do. And,

11:25

and,

11:27

actually, I'd spoken to friends of mine at Google, who were actually in the ethics team. And they had actually told management that this was misleading, and that they probably should put warning signs up, you should probably warn the people that this is not a computer. But management in their wisdom chose not to. And indeed, I played the I played the demo to my, to my wife. And she listened to the demo. I didn't tell her what it was. She said, Oh, the computers, the person answering the call, isn't it? Was it? No, no, the computers, the person making the call? It's very easy. It's I've asked a lot of people, it's about 5050, whether they get it right or not. It's very convincing. So anyway, I was answering calls about this and people's concerns about whether this was this was an acceptable thing to do to have a computer deceive you into thinking it was a human. And then I started receive calls on the same day. But another story that was breaking at the time. And this was the time of the wedding between Harry and Megan. And it was revealed that Sky News, a TV channel was going to be using some facial recognition software, so that they could automatically recognize the celebrities that were going into out of the weddings. And he didn't have any of us anyone's permission to do this have any permission. This laboratory, certainly to identify them on screen. And so again, there were journalists started to be rather concerned about issues around big brother and what facial recognition could do and whether we were being surveilled, whether people's permissions were being asked. And then it struck me because people, the phone would ring and some journalists would say, Oh, this is so and so from the ABC. And I can say, you know, Which story do you want me to comment on? Because there are multiple stories today that are breaking that people journalists ring me up and asking me about. And at that point, I realized it was a full time 24/7 job just answering the journalists questions about the latest tech fails. And secondly, it struck me in both of these cases, that

13:54

this was,

13:55

this was just bad behavior. Invading people's privacy without their permission, pretending to be someone you're not. If I knocked on your door and pretended to pretend to be someone who I someone you're not. If I knocked on your door and pretended to pretend to be someone who I wasn't. And if there was some company that was paying me to do that, you'd be rightly upset and say, Oh, this is deceptive behavior. We shouldn't allow companies behave in this deceptive way. But just because it was a computer doing it. We should be equally concerned, but didn't require us to think of anything new to worry about these sorts of issues. These were old concerns that we've always had about people being deceptive people's privacy being invaded. And so one of one of the key messages I hope you take away from my lecture today is that lots of the things that we think need to think about are actually things that we've thought about with other technologies and other settings before it's nothing. There's nothing magic here. But nevertheless, there's been a lot of concern about what a The appropriate Ethics What are the right? Moral frameworks for thinking about technologies like artificial intelligence, lots of governments who got into the act. Here's the timeline of various governments who have commissioned ethical guidelines and frameworks to do with technologies like AI, you can see Australia's on that list. And indeed, I was partly responsible, I was on one of the committees that helped draw up. Australia's ethical framework is now being implemented by a number of big companies and small companies are trialed, at least by a number of big and small companies.

15:38

So there's been

15:40

dozens of these ethical frameworks put up by governments, it's not just governments, or the big tech companies that got involved, Google has a set of ai principles, Amazon, Microsoft, you can go and read them. And of course, there's not a huge amount. There are some small differences, but there's not a huge amount of difference between what they say but they do say different things. And not just corporations, various professional bodies and non governmental organizations like the standards organizations, IEEE, an ISO who make standards. The OECD has also put up I think it's 21 ethical principles that Australia subscribe to as well. Again, I must admit some responsibility here, I sit on the IEEE and ISO committees that have helped form some of these standards.

16:32

But as I said,

16:35

a lot of this, a lot of what they say isn't very new. And that's not surprising. This isn't the first technology that has touched people's lives. And there, there are very few things that are different, I'll come to the one difference there is that most of the questions we ask are the questions we should have asked pretty much about any technology. Of course, that doesn't mean there aren't new harms being committed. Because one thing is the case of technologies like this is that they change the speed, scale and cost, you can do things. So you can now do things that's speed that you couldn't do. So facial recognition, we'll come back to that in a second. But facial recognition is a good example. It can allow us to surveil on a scale and a cost that we've never been able to do before we've we've we've had to put up with surveillance in the past, the police have stood there with video cameras and video recording, and they've looked at the tapes afterwards. That doesn't scale as well as computers. We can scale people in real time, we can scale people across the city, or as we'll see across the nation, in real time, using facial recognition software. So that does change the harms. Because of the speed, scale and cost. We can do things bigger, faster, cheaper than we ever could before. So there are and we'll talk about I spent actually the rest of the rest of the lecture talking about the new harms. But just to reiterate, the questions we ask are pretty much the questions you should ask about any technology should ask them about blockchain should ask them about 5g, you should ask them about genetic engineering, you should ask them about, about any technology that we let into our lives. And actually, I think it's worth thinking carefully. We thought carefully for 100 years about other technologists that we've let into our lives. And the place where we've probably thought most carefully for not only because it's it's hundreds of years old, but also because by its very nature, it touches people's lives, it's all about life or depositions, is medicine. That's the place where perhaps we have some of the most thorough, most carefully thought through ethical principles for thinking about how medical technologies and medical interventions should be considered and how we think about the rights and wrongs are of these. And there are four fundamental principles, four cornerstones to medical ethics, Biomedical Ethics, that I'm going to briefly mention and their facts, actually, I will argue, with one edition and one exception, make actually a pretty good place for thinking about the ethical deployment, responsible deployment of artificial islands or indeed any technology. So these principles of beneficence do good nano maleficence do no harm, autonomy, and justice and I'll go into exactly what those mean now.

19:51

So, as I said,

19:55

one of the takeaways from today's lecture is actually these these four principles plus additional. And one exception, I think actually made a really good place to think about artificial intelligence to think about the, the ethical dimensions of responsibly deploying AI. So I did say there's one edition, that one edition is what is called is a well known principle, but turns up in various other settings, especially in actually quite interestingly enough in environmental settings these days, thinking about how we're harming the planet, is the precautionary principle. And I'll talk about exactly what that is. But it's a principle that's enshrined in un law. It's been used to suggest caution, where there's uncertainty, especially causal uncertainty. I'll go into that in a second. So those five principles, I think, actually make a really good, comprehensive, it's hard to see what's missing, when you look at those, and indeed, those ethical frameworks that governments proposes ethical principles that corporations are proposed, and that standards bodies and other NGOs propose, or can be seen in terms of one of these four, five Cornerstone principles that get used in PubMed. Okay, so the first principle benefits switches. Do good. And many principles, the Google's AI principles actually explicitly say this, you should, you should look for only bringing procedures that have a overall positive impact. Not most technologies are going to be make things better for absolutely everyone. But overall, they must improve the net good of the planet or the society in which it's being used for beneficence is that closely related but subtly different. This is a very sort of utilitarian view of life, right, looking at the net benefits of things is non maleficence, which is do no harm. And that's not the same as beneficence. beneficence does allow you to do some harms as long as the goods outweigh the harms. But not malevolence, is actually no, you've got to be careful, you've got to think about the people who are going to be harmed, it's a much more egalitarian principle. And that's where things like face recognition are problematic because they do commit. And they do are plentiful examples where facial recognition gets things wrong. Facial recognition is notoriously bad at recognizing people of color, not your face recognition is notoriously worse at recognizing women than men. And it's even worse at recognizing women of color than any other group. And there are examples where people have been wrongly arrested black black men in particular, have been wrongly arrested, because they were mis identified by facial recognition software, and have ended up losing their jobs losing losing their homes losing their livelihoods. And so non maleficence would tell us that we shouldn't be careful about technologies like facial recognition, because there will be harms committed by the mistakes that the technology makes. The third principle is autonomy. And so with, think of this in the setting of medicine, autonomy is a really important principle in medicine, of respecting the autonomy of the patient. Which means that when you speak to your doctor, they try and explain what they want to do, and they get your informed consent. You're not allowed to do anything as a doctor without the informed consent of the patient, and if the patient is too young to give their informed consent of their their guardians. And you can see that that again, is important in a technological setting. That is one reason that people should be rightly upset about the duplex demo that I told you about.

24:06

Because it was being deliberately deceptive.

24:12

It was not seeking the informed consent. It didn't begin the telephone call by saying I'm Toby's computer resistant, ringing you up to book a haircut. It just pretended to be a human who deliberately tried to deceive us not to get our consent. And after all, one of the most valuable things we have is our time and if computers can waste our time, effortlessly, where are we going to be? And the fourth medical principle is, is what's called justice. This is a bit of a grab bag of various ideas. It's talks about issues of fairness, and we'll come to issues around fairness very shortly. Um, talks about whether the benefits and burdens are spread equally. And whether we respect existing laws. And as well come to work quite extensively shortly about algorithmic bias and fairness, and racial discrimination or those sorts of discrimination, sexism that we can see in algorithms that violate this principle of justice that we, that we all subscribe to. And then I said there was an additional one that's not normally mentioned in medical ethics, but I think actually, is appropriate for technologies like AI for judges, in particular, because of the way that we can scale things out. I mean, one of the, one of the exciting things about working in AI, is that you can read a research paper one year, and the next year Google's implemented as part of their search engine, and it's touching the lives of billions of people. Or Facebook can modify the algorithm machine learning algorithm for its News, the news feed. And in days, it's changing the news that billions of people are reading. And that ability to scale and have impact. Global Impact is something pretty unique technologies. It's not we've not had technologies in the past that we've been able to go out and scale so quickly, and does require us therefore to I think, be a little more careful that we are normally the technologies where if there are bugs and problems, they get found out quickly before they harm too many people. Well, the problem with these technologies that we're talking about is that we can quickly touch millions of people. And therefore the harms can be quite significant. Because even if the harms are quite small, multiplied by billions, they become quite significant. Which is why I would suggest this idea of the precautionary principle, which as I said, it's enshrined in un law, it's something that's been used to talk about protecting the environment to talk about climate change, and so on, which is that, you know, if there's a possibility that some activity or some technology is going to harm human health or harm the environment in some way, then even if we haven't worked out exactly the cause, cause and effect and exactly what how the harm is going to happen, or exactly the roots to harm is going to be, we should be precautionary, we should be careful. Even before we've established the link between lead paint and cancer, or between smoking, and cancer, or between social media and mental health, there are lots of examples I think you can come up with where we need to, especially as I said, the way that we can scale these technologies means that the quickly the harms can be magnified by the number of people we're reaching. And as I said, That's inspiring that in the Kyoto Protocol, you can find it in European law, typically, as I said, applied to the environment. But I think these are technologies where we should also perhaps consider

28:24

precaution as well. Okay, so I said there was.

28:30

So mental health is one example where these algorithms are starting to possibly how we don't know necessarily the exact causal links here. But precautionary principle would suggest we should be, we should be more rather careful about deploying these algorithms before we understand their true impacts. Okay, I said that there was one exception. And this is the exception.

28:54

autonomy.

28:56

I already mentioned autonomy when I was talking about the autonomy of patience. Now I'm going to be talking about the autonomy of how algorithms computers, that is, the one as far as I can see, having looked at the problem. There, there is one new thing this technology brings. The other technologies didn't bring in the past, this idea that machines would have some sort of autonomy, some sort of ability to act on their own, without significant oversight from humans. You get in this car, you say take me home. That's all you do. And the car makes all the other decisions. He works out the road to drive. And then it will drive me down the road. It's making decisions about obstacles, pedestrians and cyclists that it needs to avoid. It's given a huge amount of autonomy, autonomous self driving car. And that is a new thing. We've never had machines that could make decisions without much human oversight before. And the problem here is of course that The there can be consequences to these decisions cannot run someone over. And then we face the interesting ethical challenge. Well, who's going to be held accountable?

30:15

One thing is certain,

30:16

it's not the machine a machine isn't a moral being, it can't be punished, it doesn't have feelings. What can you do can turn it off, turn it back on when it doesn't care, right? Only humans can be held ultimately accountable for their actions. So ultimately, it's some humans need to be held accountable. And one of the challenges here is that there are multiple actors in play. Was it the manufacturer of the car? Or was it the was it the software program? Was it the person who sold you the car was it the person who turned the car on gave it the directions, and there are lots of people you can possibly point fingers at. And as I said, but the one thing you can't point a finger at is the car because the car is not a moral being. You can't hold it accountable for its decisions. And so that is a fresh, ethical challenge, where we have to think carefully about the choices we make and about how we deploy this sort of these sorts of technologies in a responsible way. I should warn you this picture, this picture isn't someone who's run over by the way, Mo autonomous car, that's the repair guy who's towing it did have an accident. I don't believe anyone was killed in that accident. But there have been accidents where people have been killed in autonomous cars. And I, it does trouble me how we build autonomous cars. If I told you, there was a drug company, and it was testing its drugs on the public, people like you, it didn't have any regulatory any significant regulatory oversight, it hadn't gone through ethics approval, and that actually already killed people. And they will undoubtedly going to kill more people, you'd say, well, wait a second, Toby, that that doesn't happen. Surely no. And that doesn't happen. For our companies have huge, great regulatory oversight, they have to jump through very large hoops before they can test their their products on small groups of people. And before it's allowed to be used on the wider public. But if you think about how autonomous cars are being developed, they're being driven on public roads. And people have been killed, pedestrians have already been run over by autonomous cars. And there is very little regulatory oversight, not it's not the sort of scrutiny that drug companies and you're not to this level of scrutiny that drug companies have. And you do think that's perhaps not the right way to go about it, especially because as I said, you have this accountability gap, this idea that, you know, Will, the car that's causing the accident can't be held accountable for its its actions, because it's not a moral being who we're going to hold responsible. And it does strike me as very strange, because we've done this in a responsible way previously. And again, if we look at history, the good example is aviation. So 100 years ago, when we invented flying, aviation was terribly dangerous planes used to fall out of the scale all the time, was incredibly indeed. One of the Wright Brothers was very severely injured in the very first aircraft accident, and he killed his passenger. I can't remember which of the Wright Brothers it was, but, but that's just an example of how dangerous it was right? So one of the very first passengers was killed by one of the Wright brothers. But it quickly became a very regulated industry, in which there were bodies like the Civil Aviation Authority and so on, that oversaw the development of aircraft and oversaw the safety of aircraft. And so for example, when there was an accident, there was always an independent investigation. And the lessons from that investigation were then shared with all of the industry, not just the company that built the car, but all of the players. And that doesn't happen today with autonomous cars. And that's how aircraft became so safe. aircraft flying now is the safest form of transport by various measures. And by passionate kilometers it is it easily the safest form of transport. Listened on facts. You're more likely to die in a traffic accident driving to the airport and going on a plane. Doesn't matter where the planes going. It's more dangerous. You should always when you get to the airport, you should always say relief because you've done the most dangerous part of the trip, which was Getting to the airport, getting to the airport. And now flying is much safer than the bit you've just done. And that happened. Because air craft safety is like a ratchet. Every time there's an accident, it gets investigated, people find out why the accident happened. And then a directive is issued to correct whatever went wrong to change the behaviors of the pilots to change the equipment on the plane to ensure that that accident or that type of accident never happens again. And then hopefully, that those lessons are implied, and actions will still happen. But they will have to be other types of accidents because that accidents now will be avoided. And that's where we need to get to with autonomous cars when we need to. We need to ensure that the industry shares information at the moment, there's no a sharing of information across industries, it's a very competitive field. The only information shared ironically, is the information they steal from each other. And there's been a couple of famous court cases of industrial espionage between autonomous car companies. But we need to think more carefully about to do it. So that is the one exception, autonomy. A lot of the debate, then you've seen all the interesting ethical debate you see around autonomous cars, but not surprisingly. And then the other place I've been very passionate about. Indeed, I'll be talking to the assistant Foreign Minister straight after this lecture about as a total autonomy and warfare, the way that we'll be changing how we fight war, when we hand over the killing to machines, it's a different problem for a very unique fundamental reason, which is that autonomy in cars, the cars are designed not to kill people. Occasionally, they will kill people in error. In autonomy in warfare, we're designing machines that explicitly is designed to kill people, occasionally, they will kill the wrong people. Okay, so that's the only exception. And I said, but that doesn't mean that we don't have to think about new problems and new harms, because we will break things faster, cheaper and bigger. And so I'm going to spend the rest of the time talking about how we do break things faster, cheaper and bigger. And the sorts of questions we need to know the sorts of challenging ethical questions that that that throws those up. So here's an example. Here's a heat map of crime. This is, here's this, this is Washington DC, which at one point was the second most dangerous city for being murdered in in the United States. And this heat map is telling you where crime is. And, you know, if you're running a police department, you look at a map like this, you think, Well, I don't have enough police officers to patrol every street corner, I should probably focus them on the red spots, because that's where the crime took place. And this is called predictive policing. Of course, we're not, we're not It's not Minority Report, we can't predict every individual crime, I can't say that you're about to go and rob a bank or robber corner store. But on average, averaged over a day or a month for a year. Over neighborhoods, we can make some pretty accurate predictions as to where crime took place. So it sounds like a reasonable idea. Let's concentrate our limited resources on where we expect crime to occur. Except, and there are some fundamental, moral, ethical challenges that get thrown thrown out when you start thinking about issues like this, which is that?

38:59

Well, perhaps

39:04

this, this is, you know, this is historical data. So by its very nature, it's based upon the past. If we're not careful, we're just going to perpetuate the past moving force, perhaps police officers tended to stop more black people. Most police officers are somewhat racist. And because they stopped more black people discovered more black people carrying drugs and arrested them, prosecuted them, and more black people ended up in prison. Perhaps those judges were a bit tougher on black defendants than white defendants. And so perhaps, this figure is actually somewhat distorted. It's reflecting the biases of the system in which the data was collected. And if we're not careful, we will perpetuate those biases. And then there's a there's a deep, fundamental philosophical problem And this is true not just of predictive policing. But there are so many applications of machine learning is that we're trying to predict a statistic, where we don't have ground truth. We don't know where crime took place. And it's not clear we'd ever be able to measure where crime took place, because there's lots of crime that took place, and other bits of his map, where no one reported it. No one knows that crime took place. This is only this a proxy, we have a proxy for what we are trying to predict. We have crime that took place that was reported and prosecuted. So it was a subset of the crimes that actually did take place in Washington, DC. And so often, we actually don't collect the statistic. We don't have ground truth, what we're trying to predict.

40:52

And it's not clear how you would ever collect the ground truth.

40:58

So this brings me to questions of fairness. I talked, I talked about how justice was one of those Cornerstone principles around fairness. And I want to spend a bit of time talking about fairness, because so many of the challenges turned out to be once issues of fairness, predictive policing, and was one issue first, were we being fair to black people, if more black people have been arrested in the past, to continue to perpetuate that moving forwards, and you can't talk about fairness, without talking about this, this this classic case, it's staying in policing and the justice system at all called compass that was developed by North Pole North Point cooperation in the United States. It's in use in 20 of the 52 States and United States, dues by judges and other people in the in the court system, to help work out whether someone's going to reoffend. And they use that to help decide parole decisions to help decide sentencing. So it's a computer program that tries to make them help provide a more evidence based decision. To these difficult decisions, should this defendant be given parole? Or is he going to? Was she going to commit another violent crime? While he or she is out on parole? Should society be protected from this person, even though they have yet to be convicted of the crime in which the accused? difficult decisions to make and so can if we can use a computer program that is much more evidence base? Does that not sound like a good idea? This is famous graph also, that is often used to say, Well, let's not forget what we're comparing against. We're comparing against humans making these decisions. And humans are notoriously bad at making decisions. We're full of cognitive biases. Our decision making is not at all very evidence base. This is a famous graph. from Israel, this is the probability on the y axis that's the probability

43:13

that the judge

43:17

will decide favorably to release someone they've on the dataset, they've tried to normalize against the probability that actually they committed the crime. So they've tried to make people equal weighted here. So you might ask, Well, what's on the x axis, right? So there's the probability you're gonna get get let off? What's on the x axis? What's changing your probability of being let off? Anyone? Yes, it's time of day, right. And indeed, as your blood sugar's drop, you become grumpy and angry, and you're mature, the judge is much less likely to let you off his blood sugars, good. blood sugars go up, becomes a bit more generous. as the day wears on before lunch. Again, probability drops. Clearly, we're not very good at making decisions in an unbiased way, or allowing our blood sugars to influence significantly operability and letting people off. Now, I should point out there have been various criticism made of this graph. But putting those aside, there is a certain there nevertheless remains a certain truth we know that our decision making is strongly influenced by our physical well being or our blood sugar level on the leg. So why not use a tool which is going not going to be suffering these sorts of cognitive biases that humans have? Well, first of all, you have to ask, you have to look at the design of the tool. The tool begins by asking people 100 questions and then There's a self reported questionnaire that the defendant has to fill in about themselves, which are the data points that go into this machine learning tool that are used. And you have to look at what is in this questionnaire, it asks some pretty strange questions. It asks, at what age were you when your parents separated? Now, clearly, there is if you look at the data, there is some significant correlation between people who come from a broken home, home where their parents are separated and cry. People who come from from from broken families tend to commit more crime. That is certainly the case. If you look at the data, it's undeniable. But correlation is not causation. And do we really want to be convicting people based upon them marital status of their parents? Again, strain? Interesting moral question. Here's another one. How often have you moved in the last 12 months? And again, if you there is a correlation between people who are somewhat more homeless, we've moved around between lots of places and crime. Certainly, if you look at the data, there's certainly that correlation. But but equally I had to move recently was my fault if the landlord gave me notice. Should I be less likely to be given parole just because my landlord decided he wanted to sell the house I was living in? Yeah, that doesn't seem to be very fair at all. You see, I'm appealing to principle that of justice, fairness. And here's another one, I think is a lovely one. How often do you feel bored? Presumably, there is a correlation between people who are bored, bored, and commit crime and people who are less bored who commit less crime. But do we really want boredom to be a sometimes I admit to being bored? Thank you, and UNSW it won't upgrade my Mac quite now. I just got that too. And I was terrified for it was going to do?

47:19

Well, again, I appeal to the principle of autonomy, we should respect the autonomy of UNSW academics to make the decision themselves, right. So we shouldn't just upgrade their machines without their consent.

47:33

Okay, so

47:36

there's 100 data points, they're collected this questionnaire. Some of them are things that are reasonable, like the number of prior convictions

47:46

and their age, and so on.

47:50

But people, there was big controversy when it was discovered that this program had learned from the data to be somewhat racist. It was somewhat biased against black people. And these bar charts, I think, illustrate exactly

48:10

the bias. The competitive, the competitive.

48:18

In the bottom, we're broken out into hyper steel, black and white defendants. And also, there's a graph to compare it against how well, humans human judges were doing on the same task. And you can see if you just look at accuracy, so the left hand bars, right, just look at accuracy. Looks pretty good. Seems to be about as accurate as humans are. Actually, that's pretty disappointing, right? Because, you know, I said, we were trying to be less, we're trying to do better than humans, humans should be the lower bar. But at any rate, and the fact that we haven't quite got the accuracy of humans is a little disappointing, because we're supposed to review introducing these computers to do it in a more evidence based way. So you'd hope to be more accurate, but we didn't actually do that. But we're okay. Maybe having something as accurate as humans is okay, because we're saving human effort, right. And we're not quite as accurate, but when we're in the same ballpark. And the accuracy of like defenders and white defendants, both of humans and of the computer campus looks about the same. So you say, Well, why did people get upset about campus? Why? Why was there a big controversy about this tool being racist, and that's why you look at these other two bars. And you'll see why. You see there's a significant difference in the false positives and the false negatives. And if you add the two together, they sort of balance each other out. But if you break out the false positives and the false negatives, so the false positives are the people who are forcely for dicted to reoffend who will not reoffend. So those are the people that the judge is going to say, I predict you're going to reoffend. So I'm not going to let you give you parole, I'm going to keep you locked up. Even though you might be innocent of the crime that you're, you're going to be put on trial for a while we wait for that trial, you're going to be locked up, and your liberties taken away. So that that is a real harm to those people because they're denied their liberty, when they're not going to commit a crime, while they're waiting for sent for trial. And then the false negatives, those are the people, these also a cost on society. These are the people that are going to be released, because you forcely say they are negatively going negative, they're they they're not going to commit another crime. So you can release them on on parole. And of course, while they're on parole, they committed another crime, society has harmed people, you should have kept locked up. And you could steal it the way more false negatives amongst the white defendants. Or you're more likely to say, white people who will reoffend won't release them, and they commit crimes causing harm and sanity. And on the false positive, there are black people who are not going to reoffend. Do you foresee say well, and keep locked up. So it's it's it's biased against black people in a negative sense, both under false positives and false negatives.

51:48

And, interestingly,

51:50

race wasn't what wasn't one of the points, they didn't ask people whether they were black or white. It wasn't actually the program doesn't have that data. But it does have various other data. So one of the other input fields was zip code, postcode. And as anyone who's been to knighted States know, there are many parts of the United States, certainly, urban United States where zip code is a proxy for race, or black neighborhoods and white neighborhoods. And so, the system machine learning is very good at uncovering correlations and uncover those correlations between zip code, and race and then learn to be biased on them.

52:36

Some other reasons to be upset about this tool. There are lots of ways to do better than campus or at least to be as accurate as campus. And they ran an experiment with Mechanical Turk, this website where you could just ask them to pay random people to to answer your questions. They gave them $1 For arm for making a prediction. They told them a few sentences about the defendant. And they were able to be as accurate as compass. And then another thing rather than take all these 100 features that they quit collected with that question there, they actually came up with a simple linear classifier that was more accurate, and they only use two features, the age of the person and the number of price younger you were and the more price you committed, the more likely you are to commit another crime