w8_2

Tue, Nov 22, 2022 12:52PM 49:20

SUMMARY KEYWORDS

people, fairness, drones, humans, companies, problem, algorithm, system, biases, technology, weapons, autonomous cars, decisions, robots, called, question, computers, warfare, data, false positive rate

00:00

weapons, which is what many people at Google thought they were working for company, whereas don't don't do evil? Did they really, they could be if, without their knowing they could be working on developing a software that will be going into the next product cheaper.

00:22

Regards to algorithmic decisions, do you think there's an issue about the decisions that algorithms are something inherent to algorithms that cause them to be problematic? Or is it an issue mainly in data and the gathering of data and the lack of thereof?

00:38

It can be in both bases, right? The data may have bias in it. And the algorithm may have biases in how it processes that data.

00:48

So follow us as software engineers, should we be considering both of these 5050? Or should we focus more on one problem? Well,

00:56

you got to consider the, the bias may come from either place. And that's what I equally and then it's one of the significant challenges that a lot of people haven't realized, which is that it changes. This isn't like old fashioned software, right. So if you built a word processor, and it worked, you gave it to people and you could forget, right? It was going to process words, it will continue to do the same thing. It always did. There might be some bugs in it, you can fix the bugs, but it wasn't going to start, it wasn't going to go rogue on you. with machine learning systems, they're training on constantly changing to the data that they're trained on. And there are plenty of examples of where where you have to monitor the system, because it can, it can go back to the teacher what was a fine example. But there's lots of other examples of where it's not the usual software cycle, which is okay, once it's past the unit tests, you can give it to people and you can forget, you have to think Well, is it going to is it going to go wrong? At some point, you always have to go back and retrain the model, remove any biases that may now have learned from the latest data has been trained on.

02:13

It's realistic to try to chase cautionary principle from a software perspective, where it can sometimes seem like we get one research paper and it opens up Pandora's box where someone within two months of consulting, have free access to be stigmatized. It's like people talking about dolly these days. And you start with teaching out recently, anything wrong, that kind of image generation photography, say trivially Precautionary Principle kind of have issues and application into sort of widespread adoption.

02:49

It does. So I mean, it's interesting look, so when when GPT, three first came out, the precautionary principle was applied. And it was not made public and there was a very limited access given to the system. Unfortunately, or for better, or for worse. Now, there are lots of alternatives and there are harms being committed by the systems. It's not clear that the people whose work that the data is these systems are trained on are being whether their copyright is being respected, and whether their intellectual property is being rewarded adequately, or whether that's sustainable or not. There are lots of examples in this with this, with these image systems of their racist or sexist, ageist, you ask them for a picture of doctors and they give you only men,

03:53

white men?

03:55

Right? So are they perpetuating these biases in our society? And should we tolerate that or not? And one of the challenges is that we don't know how to remove them. You know, the, the developers will say, Oh, but they're just reflecting the biases of the data which they train, which is true. If you look, if you if you search for a random picture of a doctor on the web, more likely than not, it's going to be a white male doctor. Right. But that doesn't mean we should put up with that. We should accept that.

04:27

I have two questions from online. The first one is on the vein of slowing down companies that that may be being good thing. How do you think we convince companies that slowing down is a good thing?

04:44

That's a good question. I think we have to remember that companies have social licence because we give it to them. And I think we're forgetting that the company is actually a human institution was invented. The modern day Corporation was invented In the industrial revolution to allow us to benefit from technology to allow us to benefit from the technologies that were introduced, then the steam engine, electricity, the things that that have increased our productivity increased our wealth significantly, we built a structure which allow people to take risks to invest capital, and do so in a way that was aligned with public good. And they had, you know, they have responsibilities as comm director, you have responsibilities to your shareholders. And I think, companies, we might have to think about carefully about that social contract contract that we have equally. I mean, it's it's pleasing to see companies realizing that it's actually in their long term interest to behave in a responsible way. For example, with climate change, it's companies not not governments that are doing the most activity at the moment. There are lots of companies that are moving to net zero. Microsoft is going to be net zero very shortly, not just on its current operations, but since for all of the carbon it's ever admitted since it became a company. Google is similarly Net Zero. There's there's lots of other companies, not just tech companies, that are actually leading the way in dealing with this climate emergency much more much more aggressively than governments. The standards that companies are giving themselves are actually much higher than those that countries are giving themselves. So actually, it's in their long term interest. Enlightened companies will do that.

06:49

And the other question from online is the question, Taryn? Sorry, that last question was from Henry, by the way, this question is from Tandy cannery. So similar to other industries, like aviation, where there's a monopoly, the same thing exists in tech wouldn't being more stringent on AI products also provide a way for existing mega corporations to unprovoked fairly maintain and submit their market dominance? If so, how do you combat that?

07:14

Well, yeah, there is a concentration of power. And it's not just in terms of the technology source in terms of data. Now, what can pick up tickets, Google because no one has the quality and the quantity of data that Google has, that allows it to personalize your search results. And we have to deal with monopolies. Like we always deal with monopolies because monopolies end up rent seeking, which is by antitrust. We've broken up monopolies in the past. We did that with the oil companies. We've done that with pharmaceutical companies, we've done that with the banks, we will have to do that with the tech companies. It doesn't seem to me to the consumers advantage that Facebook owns Instagram or WhatsApp. Did you know Facebook owns Instagram or WhatsApp? Right? There will be more competition among social media of if they were separate entities. Equally. Alphabet, the Google parent company, Google's made it easy to remove its dominance because it's already told you how it will break itself up by its or the parent or the or the child companies that sit within the alphabet umbrella. Right? Why is there any reason why Google search and Android need to be the same company? There's no reason at all. It just allows them to be more of a monopoly, forcing them into separate companies. And that typically is a good thing. And anyone who knows the telecommunication industry will know that there was this antitrust issue happened with the bell telecommunications, Marbella, the telecom company that were the monopoly that was in the United States, it was broken up into seven companies, and that truly liberated telecommunication in United States, which is now much better, much cheaper, much more technically advanced, because they broke the monopoly into small parts, who then had to compete against each other. And the same will have to happen in the tech industry. Well, I don't have to be forced, but

09:37

okay, we should take a break and

09:39

start again at 20 past 20 past and in the meantime, national postal reporting

10:00

just now. Okay. So if you remember the compass example I was talking about just before the break, we criticized it because it was treating black people less fairly than white people.

10:15

The

10:23

and I put up that histogram to argue why it was. The great thing about trying to get computers to make these sorts of decisions is that we can measure whether they're being fair or not. But it requires us to think well, what does fairness mean? What mathematically does fairness mean? And we quickly run into a moral minefield, which is that there are at least, last count at least 21, different mathematical definitions of fairness. And I want to show you why there are so many different ways that we can think about fairness. And we do this in terms of what's sometimes called a confusion matrix. Right? If you've got to classify something tried to make a prediction, yes or no? Is this person guilty or not guilty? Is this person predicted to be guilty or not guilty? We can write a two by two matrix, which is called the confusion matrix. And we put in that in the entries of this two by two matrix, the counts of the number of decisions it makes. So true negatives? Are those decisions that predicted to be less than true negatives? That's true, negative predictive, where you're predicted to be not guilty, where the person was actually not guilty. So we can look at our dataset and count the number of times we get the right answer, when the right answer is not guilty. True, positives are the same for for guilty, the positive decisions where we say a person is guilty, and we predicted from the data that they were going to be not guilty, we can count up the number of times we get that and we want to get most of our answers to be correct. So we won't want them to be based on that diagonal, either true negatives or true positives. But then there are the areas that the algorithm makes false positives and false negatives. False positives are those that where the person was not guilty, but we predict positively, they're going to be guilty. And the fourth negatives are those where the person is guilty. And we falsely wrongly predict that they're not going to be guilty. So we can run our classifier and our machine learning algorithm on the dataset, count up the number of times it was right, which way it was right was positively or negatively, and count up the number of times it was wrong. But those fill out this fill up this confusion matrix is two by two matrix. Okay, so that's a way of looking at the accuracy and the errors of the classifier. And now we can start thinking about whether it's fair or not. And what we do is we do one of these confusion matrix for two groups for black people or white people, for men and for women, whatever, the two groups we want to say retreating fairly or equally. And and ask, well, for example, here's one of the, here's one of the fairness measures, which is people ask that the false positive rate be equal. So the false positive rate is the number of false positives, divided by the number of true negatives and false positives. So that's the percentage of people who were not guilty or incorrectly predicted guilty.

13:36

That's one way we can

13:41

we can slice this matrix. One way we can look at the data the the entries in this two by two confusion matrix, right? Which is actually looking down the first column, right, it's looking at how accurate the first column was, right? We don't have to look at just the first column. And it's easy. That's where when we looked at the compass store, wasn't doing a very good job, right? For black people, or white people, this rate was very different. So well, let's be very fair to black people. There's another thing that it's called precision, the precision of your classifier, and we want the precision to be equal across different groups. This is the percentage who are predicted guilty who actually are guilty, right? So that's the true positives, divided by the true positives and the false positives, right. And that is the bottom looking at the bottom row of the matrix. That's just looking at numbers on the bottom row of the matrix. And indeed, if you remember the left hand bar histogram that we put up, that was the same between black black people and white people. And then D That's how NorthPoint the makers of Compass defended themselves, right? By saying, well, actually, that statistic, the bottom row was the same, relatively speaking for white people and for black people, did they optimize their classification to get that? But of course, we could look at other parts of the matrix, we could look at the second second column. Here, that's a statistical opportunity. We want all opportunity between two groups to be equal. This is essentially guilty who are incorrectly predicted not guilty? Another way of slicing this matrix? And then there's, we could look at the diagonal, right? Don't forget the diagonal is where the errors are. Are the errors equally balanced? Do we make similar is the ratio between false negatives and false positives the same? Right? You would want that to be the same across the two groups between black people and white people. Of course, you don't have to look at just one of those things that are possible you could look at to equalize and other measures could equalize the odds where you look at all four entries at the same time. So there's lots of ways you can slice that matrix up and see whether groups are treated similarly. There's other types of fairness that people think about other mathematical, there's something that you'll come across, which is called fairness through unawareness, which is, if you don't want to feature to be changing the outcome, don't include it on the input. Right, so Northwind did this. It didn't include race as one of the inputs. But the problem, of course, is that there are lots of other inputs that tend to be correlated with the one that you don't want to be glued. I mentioned zip code. So it's very hard, even if you leave it out of the system, not to infer it from something else that's left in. Now, fundamental mathematical truth is, right. So there's got these 21 definitions of fairness, you'd say, well, I'll have them all. Why not have be fair in all the possible dimensions I could possibly be? Well, unfortunately, there's a mathematical impossibility. There's a very cute mathematical theorem, which says, except in trivial circumstances, trivial circumstances are that the algorithm is the 100%, right? Or 0%, right? Same thing, right, or the two groups are literally identical, there is no way to tell them apart, they, they have the same size, and same counts of everything in them. So in those two trivial settings, where the groups are actually the same, you cannot have multiple of these measures at the same time. So for example, the false positive rate, and precision cannot mathematically both be equal. If you optimize for one, you'll make the other one different. Unless your algorithm makes no mistakes, or the two groups are literally identical.

18:09

But typically, your groups not identical.

18:14

Black people might have be poor and might therefore committed more crime. But they work items identical to the white people who tend to be somewhat richer, and have committed somewhat less crime. So you can't make these fairness measures, positive false positive rate and the position the same. So you have to make a choice. And this is where it becomes difficult for people like you, you're gonna have to make a choice, you as the software programmer have to make a choice as well. Okay, we're going to build this machine learning algorithm, it's going to be trained, I'm going to include as NorthPoint did precision as the part of my objective. So equally equalize the precision across my different groups. I'll train how to adjust the classifier. So the precision is the same. But I know that will mean that other fairness measures like the false positive rate will be different.

19:05

You have to accept that. So

19:09

ultimately, what does this tell us ultimately tells us that that fairness is a fuzzy idea that society has not well defined in hundreds or 1000s of years. Now that we're programming it, we're having to make tough decisions as to what fairness really means. It's not a new problem. It's an old problem, but it still made very concrete and explicit because we're programming it. And we see this, you know, we've seen this year on our own shores with robots that I'm not gonna say much about a robot that in the interest of time we saw this in the pandemic, knighted Kingdom where there was a wonderful time, I never expected to see people demonstrating about algorithms. So as a cause of the pandemic, they weren't able to sit school living exams. So They used algorithms to give people school leaving grades. And it all went horribly wrong. You know what I mean? The interesting thing here is the algorithm did exactly they thought very carefully about what the algorithm should do. And they said, well, we don't want grade inflation, what the same distribution of grades this year as in past years, was a reasonable idea. And so they normalize people's grades so that the grades that were given by the algorithms exactly match the distribution of grades that have been given out past years. And indeed, they did that at a school not only national wide, but they did that on a school basis, or school by school basis. So that distribution of grades in your school was exactly the same as the distribution of grades had always been in school. So it's statistically identical to the past years. But of course, that that meant that it didn't matter how hard you had worked. If no one in your school had got straight A's before, you couldn't get straight A's. It didn't reflect you know, people's attack didn't respect people's autonomy, the fact that you could go in, aced the exam to do much better than anyone in your school had ever done. Interestingly enough, the algorithms were mutant in a sense, they didn't ever change, they did exactly the same thing. They never changed. They did exactly what they were designed to do. What we hadn't thought carefully whether that was fair, was it fair to normalize the grades, like they were averaged over the society averaged over scores, that was a fair thing to do. But for the individual, it was not fair at all.

21:39

Another good example, another good example of fairness is insurance. So increasingly, we're collecting data, we can make much more evidence based decisions as to where risk is right insurance about measuring risk. And interestingly, in Europe, now, there's a new law that says the insurance rates, car insurance, health insurance, life insurance, and the cyber insurance has to be gender blind, men and women have to pay equivalent men and women have to be charged the same amount of money for their insurance. This had an interesting consequence, which, which is that women now subsidizing men is bad driving. Men have more accidents than the women is. So is a historical fact men drive worse than women men, maybe it's a consequence of testosterone. But whatever it is, men do have more accidents than women. And in the past, before this law in Europe, men had more expensive car insurance to pay for the greater number of accidents they have. But because of this decision, men's insurance rates have come down and women insurance rates have come up to me in the middle. And so women are technically subsidizing lends bad driving. So it was a decision that we made as a society, what was what was going to be fair, that insurance patient depending on your agenda, even though X rays do. So that's the end of fairness, we're going to talk about some of the other harms now. And facial recognitions is one where we see significant harms. And as I said, at the start, the harms are largely due to the fact that we can change the scale and speed that we can do surveillance. We can surveil people in real time. And if you want to understand where this might take us to, there's a system in the United in China that can scan a billion faces in a second.

23:44

In case you will,

23:46

uncertain as to what the system was going to be like they have helpfully called it Skynet.

Skynet. Skynet is the broken view to AI computer in the Terminator series, right that's trying to destroy humanity. So just to give you an idea of, but it really does change the nature of what we can do. As an example of what you can do and what you previously couldn't do. There was a case where the Chinese police picked out a wanted felon at a rock concert with 60,000 people in it. Humans can't do that human eyes cannot you cannot scan a crowd of 60,000 people and pick out a particular face. We just don't we just can't do that. But the computer can do that. And that changes the world you're in it used to be if you if you went on a demonstration of 60,000 people, you are anonymous. But now we have the technology where we can identify you, if we so choose. And we saw that with the democracy protests in in Hong Kong. What was the first thing when when the democracy protesters took over the airport, Hong Kong Airport, what was the first thing they did? They took the cameras down because that was the greatest threat to their liberty to continue to demonstrate was was the fact that they could be surveilled with these cameras. And that's assuming the technology even works properly. I mean, I already mentioned that there are black people who have been wrongly arrested because they were mis identified by facial recognition software. And there are plentiful ways we know there are plentiful ways, you can fall computer vision system, right? You write the words iPod, on a apple, and the computer vision system is very coveted is now an iPod. This funny hands like that, but but, but there's lots of easy things you can do that are potentially much more dangerous, right? You can put little stickers on this is in the real world, right? You can put little stickers on a stop sign in the real world. And it becomes a yield goes out, right or a speed limit sign, right? What's interesting is that human vision isn't, doesn't suffer these flaws. I mean, humans vision stuff has its own flaws, right? visual tricks are all optical illusions, or ways of falling the human vision system. So lots of ways you can follow the human vision system, but not like this. So the flaws in computer vision systems are completely different to the flaws in human vision systems. And of course, it's also trained on data and data, as we know, from the biases, you know, Google's trained on the web, and it's loads of biases it thinks professors are, well, some of those are perhaps true, we, we might be overrated and prejudiced and bad teachers, I assure you, we're not really overpaid. And some of us know how to solve the Rubik's Cube. But some of the biases are actually very real, very harmful, right? Climate change. If you scan the web, this is what you find out. Climate change is not real a hoax, a myth? You know, three quarters of the answers are dangerous answers. So as I said, we don't have to, we don't have to worry about these new hands that are going to fall. I want to finish with with

27:21

with a with the one new ethical challenge, which is autonomy. I talked a little about autonomous cars.

27:29

And of course,

27:31

I didn't talk about the trolley problem. And maybe, since you were invariably going to ask me questions about the trolley problem. I will talk about the trolley problem, briefly, which is one of the new moral dilemmas that autonomous cars throw up, which is that the car is driving around the corner and it has to make a split decision split second decision, does it drive into the pedestrians to stood in the road? Or does it drive into a brick wall, perhaps killing the occupants of the car. The metaphor here is this runaway runaway trolley. And you get to choose whether you whether you switch the tracks, save the five people being run over, but perhaps running over one other person on the other tracks. Now, I said it's a new problem, it's not a new problem. Every time you get in a car, you may have that problem. And you may face that moral dilemma. And indeed, when I just passed my test, I faced this very moral dilemma. I was driving down the main road, and a car pulled out in front of me. Just as the lady was walking across the pedestrian crossing with a baby. And I had to make a choice who was going to run into there was no way I could stop in time I was doing 70 miles an hour, I had to make a choice. I made the choice of driving into the car, which was the right choice because no one died. Actually, I'm not sure I made the choice. I probably just froze. But anyway, freezing was the right thing to do. The car was written off his car was written off, says it was his mistake because I was on the main road and the lady crossing to the to the island and the road was fine. So we faced that problem before. What's new, is that we have to program it in advance. Think carefully about what do we what do we program the computer to do in such a circumstance? As a human, no one had told me what to do. If you read the highway code doesn't tell you what to do. If you do the wrong thing and survived the accident. Maybe you'll be prosecuted for for manslaughter if you've done the wrong thing. But it's not explicitly written down what you're supposed to do. Now we have to think carefully what are you supposed to do? What's interesting, of course is actually I do have an April write software for autonomous cars, and you ask them about this problem. And they look at you blankly. They said, there's no subroutine, it's called the trolley problem. There is nowhere in the control loop of the algorithm that's controlling the car, where it makes these sorts of decisions, it doesn't actually understand the world well enough to make these sorts of decisions. If you look at the controller, top level controller of a car, it's drive on the Green Road. So if you've ever seen video of autonomous cars, or sensors, they're perceiving the world. And they paint the bit of the road in front of them, where they think it's safe to drive greens and drive on the safe bit of road. So that's the top level can drive and drive on the green tarmac. Or if you can't find enough green tarmac brake as hard as you possibly can. That's literally what autonomous car will do. It's not making these decisions to high kill the one pedestrian here are two pedestrians that doesn't understand the world well enough to make these sorts of choices. So it's interesting that it's captured our imagination as a moral question. Indeed, this law in Germany, that's been passed by the German lawmakers, which says, You cannot base your decisions as to who you kill in your autonomous car, on the age of the people or any of the other, just distinguishing characteristics other than the number of people. When cars, software and cars actually don't make these sorts of choices. It's actually beyond their capabilities to make these sorts of choices. And it's not clear when they will have the capability. Again, and again to end by spending a bit of time talking about one other great moral challenge that autonomy throws out, which is autonomy and warfare. And to say that, here we're talking about lethal autonomous weapons are as the media likes to call them killer robots. Just to say, it's not this guy, the Terminator robot that we're worried about. I'm not sure if we'll ever build a robot this sophisticated. But if we do that, still 10 decades, centuries away. It's much simpler, more immediate technologies that, indeed, we started to see already in Ukraine and elsewhere, that are being deployed. These sorts of things, the sorts of drones that we see today, that are increasingly autonomous. And of course, you you understand the reasons why they're increasing autonomous. One is the weakest link, and a drone is the radio link back to the person controlling it. If you can remove that radio link, it's a much more robust weapon. Also, if you're going to fly swarms of these things, humans can't fly in swarms of drones, we don't have that mental cognitive capability to fly swarms of them only computers do. So there are a number of of obvious advantages. And then there's the human advantage. The US Air Force flies a lot of drones, predator and Reaper drones. Interestingly enough, the pilots of those drones suffer Post Traumatic Stress Disorder, at the same rate as human parts, is damaging to them as it is being a real pilot and a real plane killing real people. So I've become very awakened. I've become very heavily involved in the debate around this. We would ask an open letter. That was sound by saying by 1000s of my colleagues, it was Elon Musk and Steve Wozniak, the co founder of apple and late Stephen Hawking to sign a letter which got headlines. We put out a second letter. So I learned from that first letter that you know, having people like mask on your letter tends to get you gets you in the news. So put out a second letter from from the CEOs of companies, AI and robotics companies. I'm the Sydney professor. To see I put out the second letter, and spoke at the United Nations about the challenges and the risks. In those letters, we've all wondered that would be an arms race. And you see that it said, you see that arms race happening in every theater of war, right. In the air, we see simple drones we see much more sophisticated drones like this is BA Systems to his drone it can fly across the ocean. It is an evil looking thing. It's supposed to be able to drop bombs fully autonomously. And the US Navy have this is the US Navy's first, fully autonomous ship, the sea Hunter made a Trans Pacific voyage without any human humans or bodily interesting enough when they launched the ship, they said, Oh, it's not gonna have any weapons on it. It's just gonna be used for spotting mines, identifying submarines. Now they're starting to put weapons on it. On land, you see autonomous tanks. This is Russian, MLK, 25, autonomous tank, and undersea, you see autonomous submarines. This is going to echo Voyager sizeable size of a bust, but it can travel also, again, across the Pacific entirely without human oversight. So what what are the moral ethical challenges that such weapons followed throughout? What is that these are going to be Weapons of Terror? Can you imagine what it's like to have a swarm of these drones? come at you, then indeed, that the perfect weapons for terrorists, right? Terrorists don't care, then they will commit any instructions you ask them to do? Right. So you previously if you wanted to do harm, he was doing evil, you had to find yourself an army, persuade that army to do your evil intent. Now you wouldn't, you would say go and kill all the children. And the program would do that it wouldn't question would have to brainwash the program into doing that. They will be also weapons of mass destruction. I mean, the great thing about a computer program is if you can do something once, but a loop around it, can do it a million times. Right, so I could buy a truckload of these things. 1000, these were 1000s of these drones. If one drone can kill take out one person 1000 drones could take out the first 1000 people. And I only need one programmer to do that. So it lets me industrialize, it lets me scale warfare in a way that only other weapons of mass destruction like nuclear weapons, allow us to do allow one person to take out vast numbers of other people. And they will be terribly destabilizing, right. So we already see this happening in Ukraine.

37:02

That'd be cheap. And D that makes them more worrying the nuclear weapons, right that nuclear weapons are expensive, you need a lot of technology. Need some serious physicists, you need some fissile materials and uranium or plutonium, or difficult things, you need the wealth and expertise of a nation state to build a nuclear weapon, which is helped limit their proliferation. These things you don't just need a drone, the sort of drone you can buy in your local hardware store. What will distinguish them is the software software is easy to hack. So that's again, not going to be terribly difficult. And then there's this problem of attribution. We, you don't, you're not going to know who's coming at you. And indeed, there's been a drone attack on a Russian base in Syria, where we don't know who was behind it. Take the drains down, you open them up inside and they've got Intel chips. Well, you can't work out who programs. It's not like having regular combatants come at you where you can work out where the combatants came from. I just want to end with five of the arguments that people put out when I when we talk about the challenging worlds that autonomous weapons will take us to, to actually argue for autonomous weapons, it's worth pointing out. It's not it's not a completely black or white issue. There are arguments for and against. But I think if you consider these arguments, as I haven't seen any of my colleagues that will come to the conclusion that I've come to, which is the weight of evidence is against them, not the weight of evidence is for but let me just give you the arguments. Robots will be more efficient. I think that's why you want to ban but anyway. Well, that's somewhat true and some are false. Ultimately, they probably will be more efficient. But certainly not today. If if we look at what we know about the use of drones in warfare today, the Pentagon papers that were leaked to the drone papers that were leaked, that discovered that nine out of 10 people that are being targeted with drones in Afghanistan, and pack in the borders of Pakistan, were not the intended target. That's why you still got a human in the loop. And if you remove the human there, my target would be make only nine out of 10 mistakes. But ultimately, yes, we will make them better than humans at doing the targeting. And they will be more efficient. And that's why they'll be so dangerous. I mean, as soon as humans won't be able to fight 24/7 humans won't have the speed of reflex. They will kill all the humans on the battlefield, which is something to to be worried about. The next argument, I think this is the most interesting argument. And it's worth spending some time on is that robots will be more ethical, terrible things happen in warfare. Um, you know, raping atrocities and civilians, innocent civilians get get get slaughtered. Done by humans terrified humans, often human humans. Well, if we can program computers, we can program them not to do those things.

40:20

There are two

40:23

fundamental flaws with this argument. One is that we have no idea how to program computers to abide by the rules of warfare. International matching rules are the rules under which warfare is conducted, you're not allowed to, to, to shoot people, when they're all combat outside of battle, if they put up their arms to surrender, or if they're wounded, you're not allowed to shoot about them anymore. That's part of the international rules of war. We don't know how to program computers, but there's sort of subtle ideas. I can imagine there's a future where we can 1020 30 I don't know how many years time, we probably will be able to program computers to make this sort of subtle distinctions and reason the way that humans have to reason when they're fighting what what, at the bottom now then, though, is that is that any safeguards, ethical safeguards that we can build in can be, can be removed, every computer we've ever built has been hacked by someone at some point. And so if we did remove those safeguards, then we would have these terrible, efficient, effective weapons that will be used by terrorists and rogue states to commit real harms will be used against women and children. Here's another argument, it's often people will say, Well, we could just have robots fight, we could get humans out of the battlefield of robots fight robots, Toby, which is a terribly naive idea of warfare. Warfare is not fought in some separate part of the world called the battlefield, you know, battles over here, please. Its force in towns and cities in and amongst women and children, in and amongst civilians, in and against civilians these days it's war is, is pretty much total war these days. And the sorts of people we end up in conflict with are not going to sign up to these wars and my robots against your robots. If it was that simple, we could decide war with a game of chess does, sadly, it's not like that it's going to be their robots against our civilians. A fourth argument after put forwards is, well actually can't ban something that already exists. Which is a misinterpretation of the history of disarmament, which is almost every disarm happened, the technology existed, the banning of chemical weapons, the banning of biological weapons, banning of cluster munitions, all the technologies that have ever been banned. I'll talk a little second about how successful those bands are, where technologies existed already. So we're frequently is only one technology that was bad before existed. That was the blinding laser. People were said they were going to build planning lasers, but they didn't when it was bad. But all the other technologies that we've found, are ones that existed. And then the final one, that people obviously well, yes, very good, very nice idea. In theory, it never worked. In practice, plans don't work. Again, I think we should look at history here and say, Well, yes, bands aren't 100% effective, but they do something useful. The chemical weapon ban is not 100% effective. But chemical weapons certification is not used in Syria and elsewhere. But chemical weapons are not widely used. Chemical weapons are not sold by arms companies. If you want to use a chemical weapon, you have to develop it yourself. And that has largely limited the misuse. And when chemical weapons do get used, there are international condemnation. There's headlines on the front page of The New York Times, there's resolutions on the floor of the UN, there are economic sanctions imposed, that largely limit their use. And that's what we could hope for here. We can't hope for something it's 100% effective. None of the bands are 100% effective, but you could hope for something that would actually make the world a better place. I'm pleased to tell you that things are moving forwards. And indeed at two o'clock, I'm going to speak to the assistant foreign minister to talk about that. In just six minutes time. 70 nations including Australia, have called for action on this problem with the General Assembly just two weeks ago. And I didn't have time to show the film, but you can go and watch it at home. If you type in slaughter bots into YouTube. I encourage you to watch this movie. It's eight minutes long. So I just want to end by a couple of final conclusive words. I've talked about many problems than there are many. There's not a single solution to these problems. Is there technical issues that have socio political issues? They're ultimately often about the choices we need to make as a society that the technology is putting on steroids. We certainly need ongoing research into how do we make algorithms that satisfy some of these properties better like fairness, we certainly need to encourage the public debate because he's about making, what are the choices we make with the technology that shouldn't be just Silicon Valley making choices? That should be all of us? That needs to be an educated and informed debate. So we need education as to what are the choices we need to be thinking about? And as I've hinted at, at various points, also, we need to regulate the space better. And the Sydney tech industry has been given perhaps a little too much freedom. Great. We have five minutes for questions before I have to speak to the Minister. I

45:51

have one here on line eight or later on regard to trade offs in trading various definitions of fairness, yes. Should the software engineer be the one responsible for making this decision? Or should management that is the people asking for the system itself being more responsible?

46:12

Yes, great question, Anita. I, I think it's asking too much to push that responsibility onto software developers. And whether management should be making a decision, I think the people on which the decision on which the system is going to be used will be made by making that decision, the society of people who are being a bit having to put up with the decisions that the system has made, the people who should be choice, choosing what is fair, it shouldn't be management and decide what is fair for society, society should decide what is fair for society.

46:47

Cool.

46:51

Yes, regarding the system is not as sort of a choice into some of the arguments in favor. Other French, by history, in terms of a well what we need to we need some health effects, are we going to have X number of humans, we can have X number of humans minus the number of robotic systems that are supported? Wouldn't it be more ethical to affect our lives at risk? Instead of

47:28

there is a utilitarian argument the that that you know, has has some weight. And indeed, that first lesson I talked about, we actually just called for a ban on offensive autonomous weapon systems trying to make the distinction between Well, there are a number of defensive systems. The phalanx anti missile system that sits on Australian Royal Naval Ships, is an example of a purely defensive system that's autonomous. That is saving people's lives. And it's hard to say that that's a bad thing. Interestingly enough, though, you know, disarmament diplomat, diplomats don't really like to distinction between offensive and defensive, because like about two minutes into battle, it's a bit mute as to whether you're being offensive or defensive. Maybe you find the first fast or slow, but it's that distinction between who is the offender and who's the defenders smartness, quickly lost in the heat or backwards. So it's not one that has has carried much traction, but the whole Yes. Yeah, I mean, that's a it's an offensive weapon that's used in a defensive way. And the distinction between the two is so subtle that it's not clear. I'd be happy if we banned just offensive autonomous weapon systems. But the question then is, how effective is that?

49:00

Any other quick questions from the floor before we run out of time. In that case, please join me in thanking Toby. Thanks, everybody online Cheerio.