Intro (00:00):

Morning and welcome. Thanks everyone on this cool, frosty Adelaide morning. It is lovely to see all your faces here for what is going to be a very interesting and exciting conversation. So without further ado, I'd like to introduce Ellen Broad. One's a respect to the expert in data sharing, open data and artificial intelligence. Her forthcoming book made in made by humans. The AI condition is about designing AI ethically. So we should have some very interesting elements of the conversation today. And I look forward, the session will run for about two minutes and we'll have time for questions at the end. So please think about anything you'd like to ask Ellen. So welcome. I'd like to welcome Ellen up to the stage now, if you don't like to give her a round of applause [inaudible]

Ellen Broad (00:56):

Thank you very much for having me. I should say I feel right at home, coming from Canberra, where it is frustrated every morning. So this is kind of perfect. So I'm going to talk about artificial intelligence. One of the trends we've seen in the last couple of years is a move from what I call low stakes artificial intelligence, making predictions about the films we might like to watch a Netflix, the books we might purchase on Amazon too high stakes artificial intelligence, which is making decisions about the kind of person we might be decisions about what beauty looks like. This was the first international beauty contest judged by artificial intelligence making decisions about things like your IQ predictions about the kind of job that you might have, the interactions that you might have in society. This is an Israeli company called face ception.

Ellen Broad (01:55):

We're also starting to see the use of artificial intelligence to monitor mental health issues. To try to tell us when we are experiencing the symptoms of depression and anxiety. Some of you might have seen the recent article in the guardian about professor Michael or Miquel Kosinski his work using machine learning to try to predict a person's sexuality. He also has developed applications of machine learning that purport to identify our political orientation. We're seeing the use of machine learning to absorb our social media interactions and purport to make accurate predictions about our personalities. So we're starting to see this moved from because you watched the Incredibles on Netflix. You might be interested in Kung Fu Panda two, based on the things that you say on Twitter. You are a conscientious agreeable person who is a strong team player.

Ellen Broad (02:58):

So this is the kind of trends that we're seeing before back to some of those examples. I want to talk about some of the words that I use in this presentation, but come up quite often and a really crucial one is algorithm. It's not a new word, but increasingly we hear it as something has divorced from humans, we hear things like the algorithm is making a decision. The algorithm is predicting our re-entry into society after being convicted of a criminal offense. The challenge with word like algorithm is it's what I have. I have borrowed from Marvin Minsky. He has this concept of suitcase words, words that have come to me in a wide variety of things. And algorithm has become one of these suitcase words. When sometimes we use the word algorithm where a much simpler word would suffice and be more accurate.

Ellen Broad (03:54):

Like we could say a statistical formula is calculating your credit risk or a checklist is determining whether you're at risk of re-offending. Some algorithms are not complicated at all. One of the word, one of the ways that we typically describe an algorithm, you hear this quite often in Ted talks in introductions is it's like a set of instructions. It's like a recipe or a knitting pattern. And I don't very high level. That is kind of what an algorithm is. That's why you can use words like checklist formula in itself place. But in computer science, when a computer scientist is talking about an algorithm, it typically has a more discreet meaning. An algorithm is typically a way of undertaking, a kind of common fundamental task, like ordering a list. And your computer program is laces together, algorithms that are undertaking these fundamental tasks. And that is what is your computer program.

Ellen Broad (04:56):

And this causes a lot of confusion. When we talk about algorithms in the context of artificial intelligence, because if you're a computer scientist and you're using your knitting pattern analogy, and an algorithm is not your knitting pattern, it's like your cable stitch or your must ditch. Your program is that overall set of instructions. That's going to end with a jumper at the end. And so this is where, when we start to say things like algorithms are making decisions, understanding the wide variety of things we could be talking about is really important. It could be a computer program program. It could be an Excel spreadsheet. It could be the checklist, a pilot users on an airplane to determine whether it's safe to fly. So always digging into well, what, what do you mean when you say algorithm? What does this look like is really important? The second concept that I'm going to talk a lot about is data.

Ellen Broad (05:48):

So at the moment when we talk about artificial intelligence quite often what we're talking about is machine learning which often not always, there are some applications using sparse data for machine learning, but most of the time we're talking about using massive quantities of data to make predictions, to learn trends, to make decisions. And quite often in our head, we have this idea of data as something homogenous, gushing, natural, not helped by phrases like data is the new oil. When in reality, if you're someone that works with data all the time, it can often be messy, incomplete, it can have gaps, it can be wholly unsuitable for the purpose to which you're trying to use it. It is a kind of a very human asset created by humans, whether they're using instruments like senses or whether they're collecting information from people, it has kind of the contours of the manner in which it was collected that affect what we can use it for.

Ellen Broad (06:56):

And it's really important to think about in some of the examples that are use. The third one is artificial intelligence, which is what gets everyone coming to presentations like this it's about AI. AI can mean lots of different things drones, robotics, machine learning, and natural language processing. I use it because it's like guy Steve, most of what I'm talking about though, here is machine learning. I wouldn't say that there's anything I'm superhuman about it because artificial intelligence, as I kind of keep coming back to throughout is made by humans. We make choices as developers, as engineers as your data scientists, as your data collectors, as your product owners, deciding how it is going to get applied, we decide what will be used as an input, what our applications could be used for and the trade-offs that we're prepared to take into account.

Ellen Broad (07:57):

And this has a real bearing on some of the applications that we develop. This is where issues of bias era inaccuracy come to light. The first international beauty contest that was judged by a robot identified 44 winners from entries from over 6,000 people. And nearly all of them were white, even though the entrance that put forward photos of themselves to the competition included large populations of people from India and China, but the algorithms still learnt to correlate whiteness with beauty. There were 44 overall winners, a handful of those were Asian and one was dark skinned. And that was because the data that the algorithm was trained on, the data that the program was trained on to learn, to identify beauty didn't have a lot of images of minority faces to learn from. So this is not the people submitting their photos to the contest, but how it was originally trained to identify beauty.

Ellen Broad (09:04):

And as the chief science officer said to the media at the time when the results of this became apparent, they didn't teach the machine to correlate beauty with lighter skin. It was that in the absence of a lot of images of darker skin and diverse populations in their training data, it learned to correlate whiteness with beauty, regardless. That was something that it learned on its own. And that's kind of reflective of some of the trends that we see in our own society. You know, your majority demographics can influence the way that we think about minorities. And so when we put this data into programs into machines it learns to reflect those back at us face ception has been incredibly controversial because it is trying to make decisions and predictions about people that themselves are quite hard to create an objective set of criteria for what is the definition of an academic researcher?

Ellen Broad (10:04):

What is the definition of a terrorist? How do you teach them as a machine to match these kinds of identities when you end up inevitably moving into stereotypes of what some of these kinds of people might be? Similarly, things like high IQ when I Q itself is still a contested measure of intelligence, what does it mean when we embed it in machines and then treat the decisions made by those machines as objective, we're starting to see an increasing number of applications of machine learning in the health sector. There was actually a really interesting study by university of Sydney last week, looking at a lot of these, which can often confuse ordinary symptoms of stress. So not the instance of clinical depression, but kind of an ordinary encounter with stress in day-to-day life with a clinical condition and kind of confuse people as to whether what they're experiencing is just the highs and lows of everyday life, or actually a sign of a deeper disorder because the app in your phone can't really tell the difference and this the work done by McCall's Kosinski, which I looked talk a little bit about has itself been really controversial for the choices that he made about the data to train his it's called colloquially, the Gator, G a I D R D a R.

Ellen Broad (11:36):

You know, he trained it on images of people only between 20 and 45, all white and only used the idea of gay and straight as binary's true, then train a machine to identify sexual orientation. So it kind of tried to fashion its notion of sexuality on quite a homogenous dataset and apply magic sauce. You can all do it, make sure that you then untick afterwards so that it revokes access to your Twitter and Facebook prevention credentials, but apply magic sauce. For example, on my social media data has assumed there's a 67% probability that I mail a that I am an introvert, that I'm not a team player. And I think this is partly because I mainly tweet about artificial intelligence and technology issues, and it has learned to correlate that with the characteristics of a kind of stereotypical computer programmer. So these are the kinds of, this is the world that we're living in and ethics is becoming increasingly important.

Ellen Broad (12:36):

And sometimes kind of you hear computer program is developers talk about their job, not being to make decisions about humans, they're just writing code. And actually within the industry, there have been a number of incredibly articulate well-known scientists. Who've pushed back on this for decades. A really famous one was Karen spark Jones, who this was in an interview with the British computing society in the mid two thousands talked about to be a proper professional. You need to think about the context and motive, motivation and justifications for what you're doing. The point is that there is an interaction between the context in which the program that you're developing will operate. And the programming task that you being asked to do itself. She is one of the leads kind of pioneers in natural language processing her discoveries and her research, pioneered search engines kind of it was behind Alta Vista, one of the first kind of big search engines in the early two thousands.

Ellen Broad (13:43):

She is the, one of the recipients of the ADA Lovelace metal. The British computing society recognizes pioneers in computer science, ADA Lovelace, being the first computer programmer today. She is still the only woman to have won the ADA Lovelace medal for excellence in computer science, even though it's been awarded for almost 20 years. And I think there are, hopefully there are a great number of other women pioneers in computer science that I hope get recognized soon too. And that notion that actually the decisions a computer scientist makes a developer makes about the tools and applications that they're delivering. It can have real world consequences. If you fail to think about the implications of the services and applications you build, I really famous one just from a couple of years ago, it was James Liang who was the software engineer behind the Volkswagen emissions avoiding software.

Ellen Broad (14:44):

I'm not sure, probably a lot of you remember the media stories. It was that full years Volkswagen had installed software in its vehicles that could detect the difference between an emissions test being undertaken on a vehicle and real world driving conditions. And the software would mask the emissions in real well in those testing scenarios so that the vehicles appeared to be emissions levels compliant. And in his defense, James Liang said I was just doing what my bosses told me to do. The, his defense lawyer was quoted as describing Liang as having executed a misguided loyalty to his employer. He was still the first person who was convicted in relation to the diesel cheating case. He still one of the only ones to be convicted, which I think says something about his relative lack of seniority that he's been unable to avoid prosecution.

Ellen Broad (15:43):

But it shows you the consequences in, in significant instances of failing to kind of think about what the implications are of the tools that you're being asked to build. And this comes back to things like when we talk about building AI to detect a person's sexual orientation, failing to think about in what context would that matter to decision-makers why do you need to know a person's sexual orientation? When would you desire having this in place? And you inevitably end up thinking about perhaps countries in which this is not something that's legal and accepted. It forces you to think about the consequences of what you're doing. And the AI sexuality one was a particularly glaring example, because while we talk about the ethics of that, you know, what, why do you build it? What do you think people are going to do with it?

Ellen Broad (16:41):

But that was an example where they used physionomy justifications. They used the same theories that were used a century ago to assess women's faces and determine whether they would be a good mother based on the shape of their face to determine whether dark-skinned people had low or high intelligence.

And while they said that they disavowed physionomy in the paper, I'm really were like, we are not following physionomy in that original study. And in the machine learning example, by Makala Kosinski and his fellow researchers they still kept coming back to seeing correlations between differences in facial structure and theories of kind of innate behavior. So the study is littered with terms like straight men who are more masculine and assertive gay men are more feminine. This is related to the level of hormones they're exposed to in the womb.

Ellen Broad (17:48):

But what these Google researchers proved was without even using machine learning without anything that's artificial intelligence involved, they could determine they could build a model to relatively similar levels of accuracy asking people, just superficial questions about their physical appearance. So asking them do they wear makeup? Do they have glasses? Do they have a beard and correlating that with their sexuality? Because all the algorithm was learning to do was identify superficial cosmetic trends in populations. And those trends in the images of people from their dating profiles, they were, these were images from their dating profiles. Did demonstrate a correlation between straight women wearing more makeup in their dating profiles. There was nothing innate about it. It was just picking up on kind of trends in dress, and it really disproved some of the underlying findings in that study that also talked about a lack of facial hair in gay men being about under exposure to hormones.

Ellen Broad (18:57):

And again, they just pointed out actually, when we asked a range of gay men whether they had a facial hair or not, it was just young men who didn't and because all of the faces in their studies were of young men. That's how they ended up with that kind of correlation. Not because it was reflective of any deeper kind of biological reason. It was just, they had a narrow sample of people that they were testing. So it kind of shows you the assumptions that we can make about data, the trade-offs that we make in terms of deciding what data is useful for the machine that we're building. Facial recognition has been another one that's been incredibly increasingly I should say, controversial joy while and Winnie who is also behind this spoken word poem, which I really encourage you to look at.

Ellen Broad (19:49):

I think it's brilliant. So she's a researcher with the MIT media lab. Who's done a lot of work in facial recognition. She became well-known internationally when she demonstrated that most commercial facial recognition algorithms could only identify her face when she held a white mask up over it. And this is a well-known challenge in facial recognition, which again comes back to you have minority and majority trends in a data set. So if the faces in your facial recognition, trading data, training data majority white, then inevitably your facial recognition software becomes better at recognizing light-skinned faces. It's just, it has more data to learn from, to pick up these kinds of differences. And so on, particularly on female dark-skinned faces is when facial recognition becomes least accurate. And in ain't IO a woman, she goes through faces of many famous black icons. Michelle Obama, Ida B, Wells the first black Congresswoman and Serina Williams and demonstrates the confusion that commercial facial recognition algorithms that are widely used Amazon's Microsoft's have recognizing diverse faces.

Ellen Broad (21:06):

And again, this is this is a high challenge to solve, technically because we don't have lots of additional data to teach an algorithm to learn from in that kind of context. Your so in Australia, in the Australian context, for example, where we're building a national facial recognition database using driver's license photos from every state and territory, it's going to be faces that can be in your dataset, reflect the demographics in your population. And there's not much that you can do around that. So we, we always end up talking about ethics and ethics. It takes in AI is a huge topic right now. And so I wrote this book about it, which you should all go out and buy in two weeks when it actually comes out. Not yet obviously, but I ended up becoming more and more concerned around the focus on ethics in AI primum, non nocere, no Kere, someone in here probably speaks Latin and is going to school me on my pronunciation.

Ellen Broad (22:13):

I apologize is one of the most famous principles of ethics in the world. Does anyone know what it means? Yeah. Do no harm first agreement, no harm. And quite often at the moment, when we talk about ethics in computer science, we use the Hippocratic oath as our ideal. So we say, why don't computer scientists have a Hippocratic oath? There needs to be a Hippocratic oath equivalent for the tech industry. And this is just like made me quite uneasy because the Hippocratic oath is over a thousand years old. And if you were a physician in the 15th century, 16th century, 15th century, I think is when it started being embedded in textbooks, but it's been around for centuries. And yet the medical sector is not without its history of profound, ethical scandals. We had the Tuskegee syphilis study in the 20th century, which was when researchers monitored natural progression of syphilis in a majority African-American sharecropper population for decades, even after a cure for syphilis became apparent because they wanted to understand the natural progression of the disease.

Ellen Broad (23:27):

So you had the medical profession, essentially allowing these people to become sicker and sicker and die in quite horrific ways. Using utilitarian ethics the great greatest, good for least harm, greatest good being. We now understand the progression of syphilis in, in a way that became one of the biggest medical ethics scandals of the 20th century. I think it was during Clinton's presidency that the American government eventually apologized to the victims of the Tuskegee syphilis study. And we actually see this playing out a lot in Australia. We have had a range of issues relating to the application of ethics in sectors where we quite often use greatest, good least harm to justify inflicting harm on what mostly vulnerable populations. So one of our most famous was Harry Bailey. I'm not sure if any of you remember the stories involving Harry Bailey and deep sleep therapy, but he experimented on patients with mental health issues putting them into a coma and then conducting electroconvulsive therapy in order to understand them kind of changes in the brain and a number of patients died under him.

Ellen Broad (24:46):

And it's like, these are the decisions that we make as humans about what is acceptable and unacceptable levels of harm. And this is in a sector that has had ethics as its cornerstone. For some time we still see it today. I've kind of been horrified by this unfolding story. And we still have these discussions in the medical industry at present. What are the, what are the differences in the ways that we treat Aboriginal and Torres Strait Islander patients? How do we perceive pain when expressed by different populations? We are human and we navigate ethics according to our own ideals as to what harm means, who can be exposed to harm what kinds of harm mean most? And this is why I kind of became quite uneasy about just looking at ethics in the context of artificial intelligence, because ultimately it's still humans designing these applications and we will weigh trade-offs, which is a huge part of how we design computer programs in general, let alone systems that will make predictions about humans.

Ellen Broad (25:57):

But we, we navigate these trade-offs all the time. And I don't actually think it's true that the technology sector is without ethics. For one thing, we love putting aspirational statements and statements of ethics on everything it's on walls. It's on post-it notes. One of the organizations that I worked in our meeting rooms were called integrity fearlessness, and I can't remember the third one, but you know, like we, we kind of brand ourselves with the statements. You've got the mission statements of our most famous companies. Like Google's, don't be evil. We like to uphold ourselves as a very ethical kind of organizations. Anyone who has worked in technology will notice this language wrapped around a lot of our biggest companies, but also that if you are in the industry and I look at myself in this context, a lot of people choose to work in genuinely mission-driven organizations, social impact startups.

Ellen Broad (27:06):

Non-Profits in government civic tech has real meaning. There are people who want to work on the hard business of making government work better using digital services and technology. And there are lots of people who are kind of driven by ethical ideal. There are also really famous statements of ethical principle yeah. In the industry that have been around for a long time. So a declaration of the independence of cyberspace is kind of, particularly if you were active in the industry kind of mid to late nineties, John Perry, Barlow is a declaration of the independence of cyberspace, huge. It started off kind of government. You have no sovereignty where we gather, he wrote it in response to one of the first pieces of legislation. The U S government was implementing that would curtail certain activities of internet service providers. And it was a first and foremost, a kind of separation of cyberspace from the physical world, trying to say, you know, the laws of your world do not apply in the world that we have built and it put forward its own statement of it.

Ellen Broad (28:23):

We believe that from ethics enlightened self-interest and the Commonweal out governance will emerge. And it looked at the golden rule that kind of is it do onto others, essentially what you would want done to you? As it's stands finding principal, and this was in the nineties, so quite a smart life, relatively a small community before the emergence of the big platforms that we have today. And so it was a very optimistic utopian idea of what the web would look like. And we've seen now, well today I think the conversation we're all having is, well, actually we need something more than ethics and I optimism that what is missing for a number of the kind of work that applications that we develop in the sectors that we work in is accountability. We're having these conversations at a very high level around the functions of platforms like Twitter, Facebook, Google, what role do they play now, media in our democracy and information we receive, but also at a micro level in terms of, well, what expectations do we have of organizations building AI for the medical sector, for example, or for the delivery of welfare?

Ellen Broad (29:44):

Like what, what practices should be expected? And it brings me back. It makes me think about how other sectors have emerged over the last century. So this was the pharmaceuticals sector at the turn of the 20th century. One of the most famous was Lydia [inaudible] vegetable compound, which has cured more women and any other medicine in the world. And you had these kinds of incredible statements being made about the efficacy of medicines. It's, I'm pretty sure it holds the record for the greatest number of actual cures of women's ills. This was Lydia [inaudible] vegetable compound. The really famous one was one as safe kidney and liver cure. So remedy for Bright's disease, diabetes, and all kidney liver and urinary diseases. We saw these kinds of claims being made for all every kind of pharmaceutical that you could think of California syrup of figs, perfectly safe and natural fruit laxative.

Ellen Broad (30:52):

And these were like the dominant pharmaceuticals for households advertised in every major newspaper until Samuel Hopkins Adams published an incredibly detailed expos essay of what actually went into pharmaceuticals a hundred years ago. So Lydia [inaudible] vegetable compound was almost pure alcohol. One is safe, kidney and liver Cuba, as well as being a high percentage of alcohol contained glycerin, which is used in making soaps and ingredients typically used in making fireworks and fertilizes, the California syrup of Finks, which was marketed exclusively to babies, had more alcohol and a dose than a full strength beer. And this expo's a really kicked off scrutiny of the medical sector. And, and what I think is really telling is at the moment in AI, we're kind of in a similar situation, you can advertise a product and say anything you can say. My algorithm for facial recognition is better than humans at detecting and identifying faces.

Ellen Broad (32:05):

My employment algorithm is more accurate and less bias than any human panel will ever be without any need to demonstrate how you are, how you are deciding accuracy, what limitations your system has, what data it was trained on. It's all very opaque. There's no expectations around what we should be forced to explain about the systems that are being introduced and actually being used, make decisions, but it is starting to change a New York city council has been looking at legislation prescribing open algorithms and open programs where they're being used to make decisions about city services. They just currently have a task force looking at it, but there is starting to be discussion around, well, what expectations should we have to understand how a system works, but also the limitations and the same way that you pick up your prescription.

Ellen Broad (33:05):

And it will tell you its side effects its activity, the ingredients, and what it is, is a remedy for, what should we expect to be described about AI? We've also had K areas of legislation, the general data protection regulation being one introduced by the EU that start to put expectations around the transplant errancy of systems that are making decisions about people, sorry. Yes. On one hand, I'm like saying at some point we're going to need to have a conversation at a government and policy level about accountability, but there's also a lot that we shouldn't be doing as individuals. And the first is to, of course, be curious about that. The way an automated system works, it is not magic. These things are hard fi as a person implementing a system like this, you're making lots of decisions about, well, what data can I use as an input is the model that I'm using suitable for this purpose?

Ellen Broad (34:02):

How am I auditing its results? It's human engineering. So not, yeah. Accepting the decisions made by an automated system at face value. Being curious about that, the way they work is really important, particularly when it's about individuals. Our ability to make predictions that are accurate at an Indian visual level is still very hard in comparison with making predictions about populations, making general predictions about the way a whole kind of a community of people is moving is easier than then trying to narrow that down to a prediction about a person within that community. There are these outliers ask questions. If you don't understand the way a system works, ask questions, but most importantly, demand declination. This is an increasing field in AI. I'm happy to talk about it in the break. It is challenging whether we're talking about humans, being able to explain things to other humans or machines, being able to explain the way they make decisions, but these are the kind of where we should be treating these systems.

Ellen Broad (35:11):

The way that we treat systems being introduced in other sectors, pharmaceuticals, construction, aviation. We have a lot of expectations about what good practice this looks like, but we haven't quite articulated what that is for AI in the computer science industry. I just kind of wanted to end here because a lot of the systems that we're talking about in the AI space, when we say as a system can know your sexual orientation or know your suitability for a job, it does not know it is making a prediction based on probability. It is, it cannot see the future. It is using statistics to try and make an educated guess about what the future looks like. And I just want to take it back to Hannah from the human condition is that the new always happens against the overwhelming odds of statistical laws and their probability, which for all practical purposes amounts to certainty the new, therefore always appears in the guise of a miracle. And what I think we want to avoid is the practices of statistics being treated as straight jackets and use to make decisions and therefore cutting off the possibility of miracle altogether. So thank you.