Transcript

Why ethics?

We live in a data society. In 2016, an estimated 2.5 quintillion data records were created every day. Imagine how many have been created since then.

Data, like gold or oil in the past, is a valuable commodity. Nations and corporations alike are all clamouring to use data for their own purposes with little attention paid to the impact on individuals and communities, just like gold or oil. Data is interwoven with nearly every aspect of our lives-- financial transactions, medical diagnosis, crime prevention, credit ratings, autonomous war machines, elderly care, better-managed supply chains, medical research, smart cities, climate modelling.

Data is not just an asset. It's an industry. New analytical tools and data-enabled methods make that industry complex and powerful. Can you think of an instance where someone has benefited from data use?

What about an instance where someone has been harmed? The data industry brings great benefits to society, but also new challenges and questions. Should government agencies use data analytics to identify target and penalise welfare recipients?

Should companies use predictive analytics to refuse insurance or employment because someone is likely to develop a disease? Should political parties access personal data and use it to influence voters and shape their behaviour? Should behavioural data collected through the internet of things be harnessed for other purposes? As a society, we haven't had to think about these kinds of questions before.

Technology has always influenced the way society develops, in good ways and bad. This is especially true for the last 200 years. If you think about it, for most of the last century, it was widely believed that technology could solve the world's problems, that the world's resources were infinite, and that economic growth could go on forever. The optimistic, perhaps rather naive days of the 20th century, are long past us. There is increasing evidence that those long-held worldviews are no longer valid.

There are new questions to be answered. Corporate and political influence has increased in its scope and effects. This has led to an increased focus on the responsibilities that corporations and institutions have to the societies in which they operate. It's particularly important when it comes to the data industry, which reaches every realm of people's lives. The values and habits of thinking that may have applied in the past might not apply nowadays.

Data use has ethical consequences. It can promote or diminish people's rights and dignity. Real people are experiencing the real effects of data science in their everyday lives, yet individuals don't usually own their data.

And they have little say in how it's used. So we need to consider the duties and responsibilities that arise when data is collected and used. Let's think about the use of algorithms and machine learning.

We live in a world where decisions and choices that used to be made by humans are now being made by algorithms that decide how data should be interpreted and what action should be taken. We have high expectations and generally trust algorithms when we automate data mining and interpretation, but think about how these algorithms are made and used by individuals and processes that may be as much about trial and error and tinkering as careful planning. The way data is processed and packaged is essentially a black box.

It is inherently difficult to make those processes transparent to the people they affect. It is inscrutable. Humans are pretty good at identifying and correcting our own mistakes.

So when a person analyses data, they apply that inherent skill to the process. Algorithms enable generally reliable decisions based on complex rules to be made, but these decisions are also subjective and not necessarily correct. Machine learning multiplies the speed of analysis and decision making, but it also multiplies the speed and effect of any mistakes. Algorithms can and do make type 1 false-positive and type 2 false-negative category errors, which affect people's lives. They are often built on basic assumptions and so reflect the values of the programmers.

What would you do if you were labeled as a future criminal when you are not one? Data ethics means bringing the people who are impacted by the data industry into the front of our minds. It means being human-centered. We need to think of algorithms as more than just mathematical constructs.

There is a social element to the way algorithms are developed and used. Data selection, analysis and visualisation, all depend on decisions that reflect a particular set of values and priorities. Data can be mobilized and subverted beyond its original purpose.

It can also be given different meanings and used as evidence to support specific values, ideas, or knowledge claims. We know now, for example, that neuroscientists began with assumptions about gender when selecting and visualizing gender-based brain data. That data and the way it was presented reflected these inbuilt values and assumptions.

Technologies such as machine learning and predictive analytics might be seen as ethically neutral, but every technology has embedded values and assumptions. For example, if you collect employee data or carry out surveillance on your employees, it might be because you think employees can't be trusted or that close monitoring will make them work harder. These views reflect a particular set of assumptions about human nature and values around what is important. They are not facts. So all data collection and use is socially embedded.

As we encounter more of the unintended consequences of data decisions, we're recognising that organisations and individuals that collect and use data have obligations beyond economic or political outcomes. This isn't an either or situation where we have to choose between economic outcomes and wider social responsibility. It's a call for both. Ultimately, we must consider the question, in what kind of society do we want to live?