

# The ethics of algorithmic decision making

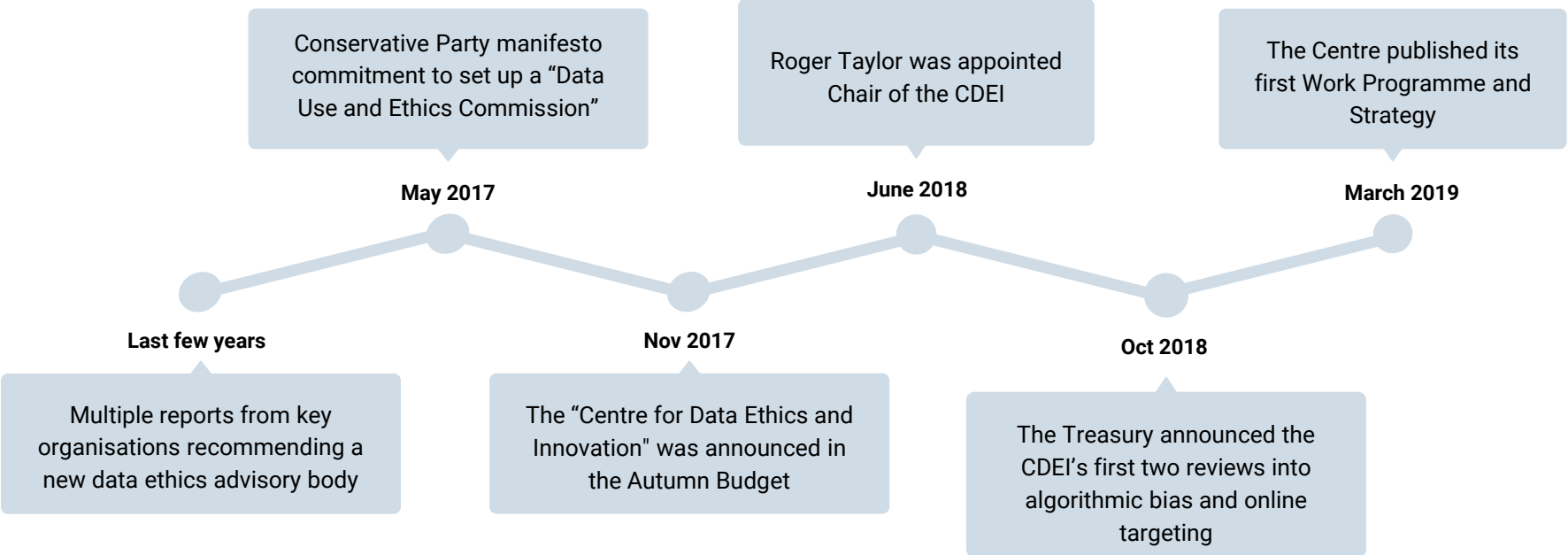
Jennifer Boon, Senior Policy  
Adviser

23<sup>rd</sup> May 2019

# Who we are

- The Centre for Data Ethics and Innovation (CDEI) is an independent advisory body tasked by the UK government to investigate and advise on how we maximise the benefits of AI and data-driven technologies
- The first body of its kind to be established anywhere in the world
- The government is committed to establishing the Centre on a statutory footing after its initial phase of operation

# How we got here



“ It is crucial that the public have confidence artificial intelligence is being used to improve people’s lives and we have the right expertise and framework in place to maximise its potential

- Rt Hon Jeremy Wright MP

# What do we mean by algorithmic decision making?

- The breadth of what can be considered algorithmic decision making is huge and these approaches are not new.
- The change lies in the kind of technology available and the scale and quality of the data which underpins it.

# What do we mean by ethics?

- “Moral principles that govern a person's behaviour or the conducting of an activity.”

Oxford English Dictionary

- "In civilized life, law floats in a sea of ethics."

Earl Warren

- How do we define what is ethical and, once defined, how do we govern for those ethics?

**What questions  
should we be asking  
ourselves about an  
algorithm?**

# Does it work?

- Efficacy is generally a necessary, but not sufficient condition for algorithms to be ethical
- What do we mean by efficacy and what level of error rate is tolerable?
- If performance varies, how does it vary?



# How does it compare to the status quo?

- It may work, but does it work better than what we have?
- If it does not work better, what are the reasons for adoption?
- Do we understand the status quo well enough to make this judgement?

# Is it biased?

- How are we defining bias and fairness?
- How does this relate to underlying societal biases?
- Where are these biases coming from?
- What new biases might emerge from the use of AI and ML?

# How is it being used?

- Are algorithms being used to make decisions or are they working in a mixed model with humans?
- What context are they being used in?
- What happens before and after the involvement of the algorithm?

# Is it explainable?

- Do we understand the recommendations and decisions which algorithms make?
- Do we need to?
- How do these standards relate to the ones we apply to human decision makers?

# What are the trade offs?

- We cannot optimise for everything, so what do we value most?
- Are the accuracy and efficiency gains worth it, given potential losses in explainability?
- Are we being explicit about the trade offs involved?

# Who owns the algorithm?

- Is it appropriate for sensitive algorithms to be commercially confidential?
- If it isn't, how can the right level of skills and investment be attracted to the sector?
- Is there a difference between public

# Who is accountable?

- Who is responsible for a decision taken entirely or in part by an algorithm?
- What are the consequences when things go wrong?
- Where does this leave human autonomy?

# What is the CDEI doing?

- We are running a review into bias in algorithmic decision making, to conclude in Spring 2019.
- We are taking a sector specific approach, focused on financial services, crime and justice, local government and recruitment.
- Visit our [website](#) for more information and to



# Thanks!

## **Any questions?**

You can find us at:

- @CDEIUK
- cdei@cdei.gov.uk