

# The Alan Turing Institute

See the full story and others like it at [turing.ac.uk/partners/impact-stories](https://turing.ac.uk/partners/impact-stories)

## Impact story

### Fairer algorithm-led decisions

Turing researchers from diverse fields have produced a new way of approaching fairness in algorithm-led decisions, by looking at the causes of certain factors that can sometimes result in biased decision-making.

- Algorithms are increasingly assisting in life-changing decisions, but if they're trained with data biased against certain demographic groups, they can produce biased decisions.
- Turing researchers have developed a framework using 'causal methods' which aims to ensure algorithmic fairness by taking into account different social biases and compensating for them effectively.
- A decision is fair towards an individual if the outcome is the same in reality as it would be in a 'counterfactual' world, in which the individual belongs to a different demographic.
- Any differences in variables used in an algorithm identified as being caused by sensitive factors, like race, need to be cancelled out.
- The framework calls for the need for decision-making algorithms to be designed with the input of expert knowledge about the situations the algorithms are being used in.
- 'Counterfactual Fairness' paper published in March 2017, generating significant interest.
- Team are working and engaging with policy-makers, lawyers, and investigative journalists, whilst refining and improving their methods.

#### Impact

- Set of technical guidelines for practitioners
- Influenced many leading researchers to use causal methods
- Helping identify racial biases in automated US parole procedures
- Potential to have wide-reaching, meaningful effect on everyday life

**“This work is the first to show us how we might use causality to uncover injustices in algorithms”**

Simon DeDeo, Assistant Professor in Social and Decision Sciences at Carnegie Mellon University



Research Fellow Matt Kusner presenting at machine learning conference NIPS in Dec 2017