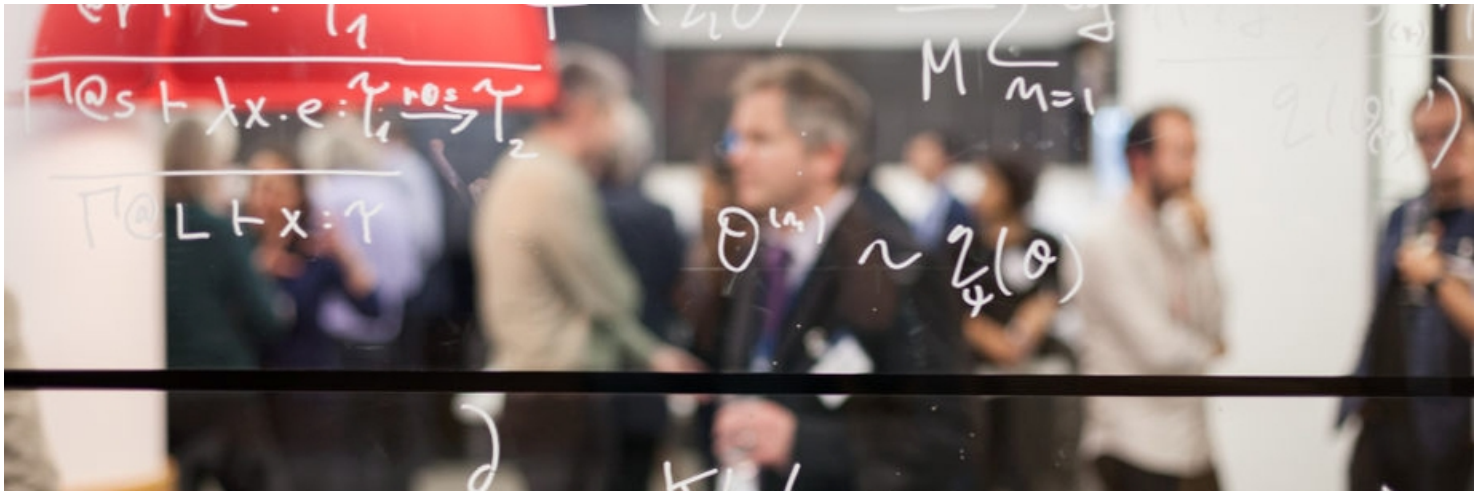


Data Science: a new discipline to change the world

 www.epsrc.ac.uk/blog/datascience/



Data science - the new kid on the block

I have re-badged myself several times in my research career: mathematician, theoretical physicist, economist (of sorts), geographer, city planner, complexity scientist, and now data scientist. This is partly personal idiosyncrasy but also a reflection of how new interdisciplinary research challenges emerge. I now have the privilege of being the Chief Executive of [The Alan Turing Institute](#) - the national centre for data science. 'Data science' is the new kid on the block. How come?

First, there is an enormous amount of new 'big' data; second, this has had a powerful impact on all the sciences; and thirdly, on society, the economy and our way of life. Data science represents these combinations. The data comes from wide-spread digitisation combined with the 'open data' initiatives of government and extensive deployment of sensors and devices such as mobile phones. This generates huge research opportunities.

In broad terms, data science has two main branches. First, what can we do with the data? Applications of statistics and machine learning fall under this branch. Second, how can we transform existing science with this data and these methods? Much of the second is rooted in mathematics. To make this work in practice, there is a time-consuming first step: making the data useable by combining different sources in different formats. This is known as 'data wrangling', which coincidentally is the subject of a new Turing research project to speed up this time-consuming process. The whole field is driven by the power of the computer, and computer science. Understanding the effects of data on society, and the ethical questions it provokes, is led by the social sciences.

All of this combines in the idea of artificial intelligence, or AI. While the 'machine' has not yet passed the 'Turing test' and cannot compete with humans in thought, in many applications AI and data science now support human decision making. The current buzz phrase for this is 'augmented intelligence'.

Cross-disciplinary potential

I can illustrate the research potential of data science through two examples, the first from my own field of urban research; the second from medicine - with recent AI research in this field learned, no doubt imperfectly, from my Turing colleague Mihaela van der Schaar.

There is a long history of developing mathematical and computer models of cities. Data arrives very slowly for model calibration - the census, for example, is critical. A combination of open government data and real-time flows from

mobile phones and social media networks has changed this situation: real-time calibration is now possible. This potentially transforms both the science and its application in city planning. Machine learning complements, and potentially integrates with, the models. Data science in this case adds to an existing deep knowledge base.

Medical diagnosis is also underpinned by existing knowledge - physiology, cell and molecular biology for example. It is a skilled business, interpreting symptoms and tests. This can be enhanced through data science techniques - beginning with advances in imaging and visualisation and then the application of machine learning to the variety of evidence available. The clinician can add his or her own judgement. Treatment plans follow. At this point, something really new kicks in. 'Live' data on patients, including their responses to treatment, becomes available. This data can be combined with personal data to derive clusters of 'like' patients, enabling the exploration of the effectiveness of different treatment plans for different types of patients. This combination of data science techniques and human decision making is an excellent example of augmented intelligence. This opens the way to personalised intelligent medicine, which is set to have a transformative effect on healthcare ([for those interested in finding out more, reserve a place for Mihaela van der Schaar's Turing Lecture on 4 May](#)).

An exciting new agenda

These kinds of developments of data science, and the associated applications, are possible in almost all sectors of industry. It is the role of the Alan Turing Institute to explore both the fundamental science underpinnings, and the potential applications, of data science across this wide landscape.

We currently work in fields as diverse as digital engineering, defence and security, computer technology and finance as well as cities and health. This range will expand as this very new Institute grows. We will work with and through universities and with commercial, public and third sector partners, to generate and develop the fruits of data science. This is a challenging agenda but a hugely exciting one.

Author

In the following table, contact information relevant to the page. The first column is for visual reference only. Data is in the right column.