Big issues

The Alan Turing Institute is bringing a different approach to some of the issues associated with Big Data. By **Graham Pitcher**.

In June 2013, Sir Mark Walport and Professor Dame Nancy Rothwell, co chairs of the Council for Science and Technology, sent a letter to then Prime Minister David Cameron entitled 'The Age of Algorithms'. The letter contained eight recommendations, the sixth of which was the establishment of a National Centre to promote advanced research and translational work in algorithms and the application of data science. "This could fittingly be named the 'Alan Turing Centre'," they noted.

'Big Data' also featured in the 'Eight Great Technologies' speech made in 2013 by ex Industry minister David Willetts in which he outlined areas where Government investment could support the development of innovative technologies and strengthen the UK's competitive advantage.

Picking up on this theme, the Alan Turing Institute was founded in 2015, looking to research some of the big challenges in data science.

So what is 'Big Data'? "It's a buzz phrase with lots of layers," said Dr Anthony Lee, one of the Institute's strategic programme directors. "It refers to the huge amount of data available as a result of data collection and digitalisation in all sectors. And everyone is recognising the importance of data to industry and the economy, amongst other things."

Why the focus on Big Data? "In some respects, Big Data has been around for a decade," he said. "But lately, industry and scientists have realised how useful the data around us is. Before, we would try to learn more about the world around us using highly focused statistical experiments, for example," Dr Lee explained. "Big Data has changed the questions we can ask; we can improve our understanding of society based on the analysis of data, while companies can build better products.

"The quality and quantity of data and information we are dealing with nowadays is very heterogeneous. Companies collect data about their customers and can extract valuable commercial benefit from it. Similarly, we may broaden our knowledge of the universe by analysing information collated by space telescopes that are either directed at carefully chosen regions of space or designed to cover a wide area."

And Big Data is big; the term is a representation of the fact that, during the last 30 years, the amount of data generated per year has increased by a factor of 10 every two years.

Dealing with Big Data entails bringing together people with a range of skills. "There is a range of inter-related themes," Dr Lee noted. "Machine learning is an important part of data science and one way of tackling the huge data sets. In the past, artificial intelligence was more about deterministic logic, while ML is more about probabilistic reasoning."

But he admits it's not always easy getting these people to work together. "We have created a strong interdisciplinary environment in which leading academics can collaborate to solve problems which they wouldn't have been able to solve on their own."

And the Institute is bringing together a diversity of people, with

The Alan Turing Institute is bringing together a diversity of people, with pure mathematicians, classical statisticians, computer scientists, social scientists and software engineers contributing to algorithms and hardware.

pure mathematicians, classical statisticians, computer scientists, social scientists and software engineers contributing to algorithms and hardware. "We really need these people to work together if they are to solve the challenges."

A different point of view

The Institute's mission as the UK's national institute for data science is to bring a different perspective on solving the challenges posed by Big Data. "It's world class fundamental data science," Dr Lee pointed out, "as well as applied research. It's not only somewhere that people can start to understand the shape of problems surrounding data, but also a place where the next generation of leading data scientists can be trained."

Alongside the Engineering and Physical Sciences Research Council, the Institute has five partner universities – Cambridge, Edinburgh, Oxford, University College London and Warwick – and four strategic partners – Intel, Lloyds Register Foundation, GCHQ and the Ministry of Defence, and HSBC.

Dr Lee, a computational statistician in the Department of Statistics at Warwick, is strategic programme director of the partnership between the Alan Turing Institute and Intel.

"Our university partners are complementary," Dr Lee noted. "They all have mathematics and computer science departments with global reputations and each has specific

strengths. It's important that we can draw on this diverse talent base when we're looking at problems.

"The UK enjoys a special position in terms of maths and there's a huge strength in statistics. Alongside that, its engineering strength dates back to the industrial revolution."

One of the issues with Big Data is that, as the name suggests, the problems are also big. So identifying which challenges to pursue can in itself be a problem

"We instigated a huge scoping process," he explained. "More than 1000 people contributed to this process, which ended up with 100 research proposals being made.

"This allowed the Institute to create a matrix of areas of interest to researchers. The matrix blends horizontal themes with vertical industry sectors. And this is one of the interesting things; the cross cutting nature of the research being embarked on here provides challenges for everyone."

Amongst the themes being explored are performance failures in large computer clusters, understanding why they happen and how to mitigate them. "There's also data centric engineering and smart cities, as well as progress on secure cloud computing."

With the focus on maths and computer science, observers might be convinced that the Institute is software oriented. But Dr Lee disagrees. "It's about everything to do with Big Data and it's why we

"Big data is almost a revolution and the important thing is that it underpins many aspects of the economy."
Dr Anthony Lee

A matrix of research interests blends horizontal themes with vertical industry sectors, highlighting the cross cutting nature of the Institute's work

have horizontal and vertical focuses. It's not only about algorithms, it's also about the systems on which they run; we'll even be looking at chip design and network interconnects.

"Deep learning is a good example; you need to create a network, then train it. You need mathematical modelling, algorithms, hardware and software skills to solve the problems."

Pointing out the close relationship between software and hardware, Dr Lee noted that Turing himself was a mathematician, but was involved with the early modelling of computers and had insights into data analysis. "Similarly," he added, "von Neumann was a mathematician, but developed a computer architecture."

Although established in 2015, the Institute's full research programme began in October 2016. "We've made good progress on our initial focus areas," he said. "Like all good scientists, we're solving challenging problems, but ones that can be solved.

"But progress is not as difficult as some might think. While all academics have core research interests, some can see the value in collaboration and will recognise they can achieve things they couldn't before."

While the Institute is endeavouring to 'hit the ground running' and solve some immediate problems, Dr Lee said it was still important to keep an eye on the 'long game', where fundamental work might need to be done.

So is the UK the leading player in Big Data, as Willetts hoped for? "The UK is a world leader," Dr Lee agreed. "That's not to say other countries aren't paying attention to the topic.

"Big data is almost a revolution," he concluded, "and the important thing is that it underpins many aspects of the economy. Already, there has been a large scale uptake in the use of Big Data in decision making."

engineering	technology	defence and security	smart cities	financial services	health and well-being
mathematical representations					
inference and learning					
systems and platforms					
understanding human behaviour					