



Discover #OxfordAI



The University of Oxford Guide
to Artificial Intelligence

Welcome

By Chas Bountra

Ask our academics why Oxford is at the forefront of AI research and they'll invariably say one of two things: opportunity for collaboration, and diversity of thought.

Not only do our researchers come from all over the world and from a variety of backgrounds, they're also working across the full spectrum of the transformative field we call artificial intelligence.

And amid the undoubted hype, these technologies genuinely will be transformative.

In this publication you'll read about how AI and machine learning techniques are allowing clinicians to diagnose more accurately conditions such as heart disease; how the banking and finance sectors are being revolutionised by algorithms; how we're moving towards a world in which vehicles are able to drive themselves.

You'll also read about the fundamental scientific research underpinning these world-changing applications – research carried out by mathematicians, statisticians, computer scientists and information engineers.

Finally, you'll hear the views of key voices in the ethical, social and legal debates that inevitably arise alongside rapid technological advancement. How do we know the algorithms making decisions about our lives aren't biased? What is the likely impact of automation on jobs? Will we ever see the day when machines can truly think like humans?

And this selection only scratches the surface of Oxford's work in these areas.

Working across disciplines, together with industry and government, with funders, third sector partners and colleagues at other universities, Oxford's world-leading researchers are perfectly placed to tackle the challenges and exploit the opportunities of the AI revolution.

Discover **#OxfordAI** here.



Chas Bountra
Pro-Vice-Chancellor
(Innovation), and Professor
of Translational Medicine,
Nuffield Department
of Medicine

Contents

INTRODUCING AI

The two routes to AI

PAGE 5



Analysing the shape of data

PAGE 6



Automating science

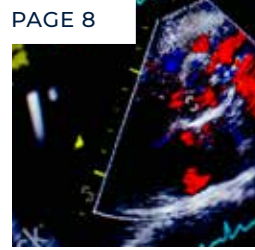
PAGE 7



APPLICATIONS

Making healthcare smarter

PAGE 8



A game changer for drug discovery

PAGE 10



Biomedical intelligence

PAGE 12



Diagnosing heart disease

PAGE 13



Autonomous vehicles

PAGE 15



Transforming the finance sector

PAGE 16



Saving the planet

PAGE 18



ETHICS, SOCIETY AND GOVERNANCE

Robots thinking fast and slow

PAGE 21



Automation and the future of work

PAGE 22



Think about ethics now

PAGE 25



Outwitting the 'junk news' nation

PAGE 26



A flourishing future

PAGE 27



The need for a legal framework

PAGE 29



Are algorithms biased?

PAGE 30



The Futuremakers podcast

PAGE 31



For further information on Oxford's AI research or working with Oxford, visit ox.ac.uk/ai or contact public.affairs@admin.ox.ac.uk



The two routes to artificial intelligence

By Mike Wooldridge

An early English to Russian translation program is said to have translated the sentence 'The spirit was willing but the flesh was weak' as 'The vodka was good but the meat was bad'.

AI is hard – I hope we can all agree about that. AI is hard for many reasons, but one of the most important is that it is a subject which is hard to approach directly.

We have evidence that intelligent behaviour is possible – we provide that evidence – but the processes that lead to intelligent behaviour are hidden from view, inside our brains. We can't examine these processes directly, and so when we try to create intelligent behaviour, we have to start from a blank slate.

So, how do AI researchers go about building systems capable of intelligent behaviour?

There are basically two types of approach, and one of these has been shown to be dramatically successful over recent years.

Let's suppose we want to write a program that can translate texts from English to French. Not very long ago, programs that could do this competently were firmly in the realm of science fiction, and progress in automated translation was so slow that it was something of a cruel inside joke for the AI community.

Famously an early English to Russian translation program is said to have translated the sentence 'The spirit was willing but the flesh was weak' as 'The vodka was good but the meat was bad'. Whether or not the story is true (it isn't), it has an inner truth: machine translation programs were plagued with problems, routinely making blunders in translation that a child would not make.

The main approach to machine translation, which was the dominant approach until this century, was what we might call model-based. With this approach, what we do is try to come up with a model of the behaviour we are trying to reproduce, and to give that model to a computer so that it can use it directly. For English to French translation, the models in question would be models of the English and French languages.

First, we would define the structure of sentences (technically, the 'syntax' – what makes a grammatically acceptable English and French sentence and text). We then use that syntax to understand the structure of the text for translation, and hopefully from that we can derive the meaning of the text (the 'semantics'). Once we have that meaning, we can again go back to our model of the target language and construct a corresponding text from the meaning.

This approach to natural language understanding requires us to be able to come up

with rules defining the grammar, how to extract the meaning from the structure of the text, and then how to generate a text from the meaning. So, researchers busily worked on all of these problems – for decades. Ever more elaborate ways of capturing text structure and meaning were developed, and there was progress. But translation using these approaches never achieved human-level or anything like it.

The problem is, human languages are complicated and messy – they simply resist precise attempts to define their syntax and semantics, and are so full of subtleties, quirks and exceptions that they are seemingly impossible to nail down.

In the 1990s, another idea began to gain prominence, called statistical machine translation. Remarkably, with this approach there is no attempt to construct any kind of model or understanding of the language in question. Instead, what we do is start with a large number of examples of what we are trying to do (translated texts), and we use statistical methods to learn the probability of particular translations. The basic maths behind this approach is simple, but to make the approach work in practice required lots of data, and lots of processing time to compute the statistical associations.

Statistical machine translation achieved remarkable successes very quickly, and the field was turned on its head. And the same ideas began to be applied in other areas. Other learning techniques were investigated and found to work – one of them being deep learning, which is the hub of all the present excitement about AI.

So there, in a nutshell, are the two basic approaches to AI. With the first, you aim to create a model of the thing you are trying to achieve, and give that model to a computer. This has the advantage of being transparent – we can look at the model and see what is going on. But for complex, real-world problems, coming up with a practically useful model may be impossible.

With the second, you don't worry about a model – you simply throw lots of examples of the behaviour you are trying to create at a machine learning program, and hope that the program learns the right thing to do. And

the exciting thing is, at present, there is a lot of progress with this approach. The very big disadvantage of this approach is that it is opaque. Your program may learn to do its task better than a human, but it can't tell you how it does it, and ultimately the expertise of the program is encoded in (very long) lists of numbers – and nobody has any idea how to tell what those numbers mean.

This conundrum is at the heart of contemporary AI. We can have transparency or competency, but at present it seems we can't have both. And that is one of the biggest challenges for AI research today.



Mike Wooldridge
Professor of Computer Science,
Department of Computer Science

Analysing the shape of data

By Heather Harrington and Ulrike Tillmann

Datasets such as point clouds, networks or images that are generated often have vital information encoded in the shape of complex features.

As the world is increasingly overwhelmed with data, the simple question arises: where do we turn to find new approaches to extract information from it?

Not only do we have data on an unprecedented scale, much of this data is complex, multiscale, incomplete or noisy. Advances in statistical methodologies, combined with powerful computers, have provided new avenues to harness the information held in datasets – but is it enough? Where can we find the new ideas that will allow us to build upon existing techniques and unlock vital information from vast datasets?

It is a smart move to look towards mathematics, and topology is a branch of mathematics that is teeming with exciting possibilities. Topology studies the shape of data from a fundamental perspective that is not limited by classical geometric constraints. Our newly founded Centre for Topological Data Analysis will develop powerful data analysis tools based on topology to grasp the geometric essence of complex and large data collections.

Datasets such as point clouds, networks or images that are generated often have vital information encoded in the shape of complex features that are not accessible through traditional techniques of statistics and machine learning. Topological data analysis (TDA) offers an exciting new way – in combination with AI techniques such as machine learning – to characterise and quantify the shape or structure of data with unrestricted complexity and generality.

Once we have computed a topological summary of a dataset obtained from the topological data analysis, we need to interpret this information. The topological summaries we obtain are not elements of a metric space and therefore are not directly suited to traditional statistical methods. A very active area of research is the development of approaches that are suitable for comparing and classifying output

from TDA. These approaches can be broadly divided into vectorisation methods, which build an explicit feature map, or kernel-based methods, which build kernels on the topological summary. Once such a map is constructed, either explicitly or implicitly, these frameworks are amenable to combining TDA with machine learning.

A wide variety of applications have already benefited from TDA, including image processing, text analysis, chemistry, electromagnetism, materials science, fundamental physics, neuroscience and medicine. One particular problem we are tackling is to characterise the shape of blood vessels as a tumour develops, combining TDA with data provided by Oxford University's CRUK/MRC Institute for Radiation Oncology. Understanding this complex and dynamic process has hitherto not been possible with standard statistical techniques, and our preliminary analyses using TDA suggest that it can provide a multi-scale, quantified description of the data. Such descriptions of tumour blood vessels will be an important component of future diagnosis and treatment tools.

It is our vision to build a two-way bridge between data users and scientists so that topological ideas and tools can flow between testing and applications, and research and development. Our multidisciplinary team of mathematicians, statisticians, computer scientists and practitioners from industry has the necessary breadth and depth of experience and expertise to drive the development and application of TDA towards the solution of real-world problems. We have partnered with leaders in academia and industry to create a pathway for the joint development of the theory and algorithms that will enable TDA to become widely accessible – and indispensable in data science.



Heather Harrington

Royal Society University Research Fellow in Applied Mathematics, Oxford Mathematical Institute, and Co-Director, UK Centre for Topological Data Analysis



Ulrike Tillmann

Professor of Mathematics, Oxford Mathematical Institute, and Co-Director, UK Centre for Topological Data Analysis

Automating science

By Stephen Roberts

We live in an era of big data. The generation of terabytes' worth of data is now ubiquitous in the modern world. In many scientific disciplines, the ability to cheaply, efficiently and rapidly record data allows experiments themselves to become a sophisticated acquisition exercise.

We are already at the point at which AI systems can infer such things as conservation properties (such as the conservation of energy and momentum) and propose underlying 'laws', given only data.

Science, namely the construction of deep understanding from observations of the world around us, can then be performed in the data. For many years this has meant that teams of scientists, augmented by computers, have been able to extract meaning from data – making an intimate bridge between science and data science. More recently, the sheer size, dimensionality and rate of scientific data has become so vast that increasing reliance on automation and intelligent systems has become prevalent. Algorithms can scour data at scales beyond human capability, finding interesting new phenomena and helping the discovery process.

The physical sciences have many examples of vast-scale algorithmic science projects. When it comes fully online, the Square Kilometre Array, a radio telescope network currently under construction in Australia and South Africa, will generate more data than the entire global internet traffic – and is already streaming data at almost a terabyte per second. The Large Hadron Collider at CERN discovered the elusive Higgs boson in data streams that were produced at a rate of gigabytes per second. Meteorologists and seismologists routinely work with complex global sensor networks that generate vast datasets, all differing in the type, quantity and quality of data produced.

Nor are the problems confined to the volumes of data now produced. The signal-to-noise ratio is often very low, and data may only provide biased estimates of desired quantities. Data is often incomplete, which complicates the extraction of automated meaning. Finally, we must always ask if the data and algorithm combination is able to answer the research question posed, and which combination of data and algorithm is the most valuable given the scientific objectives.

Addressing the issue of what data and which algorithm takes us to the issues of intelligent selection of experiments, models and methods, both to acquire new data and also to shed new light on old data. All these processes can be, and are, automated.

The concept of optimal experimental design may be old, but modern equivalents, particularly work on automated machine learning, bring intelligence into the way data and algorithms are chosen so as to maximise the informativeness gained. This can also take into account the costs (which may include, for example, economic costs, hardware and memory limitations, and time) associated with data recording and computation, enabling efficient, optimal

experimentation to be performed with a given budget.

The laws of science are compressed, elegant representations that offer insight into the functioning of the universe around us. They are, ultimately, developed by logical (mathematical) formulation and through empirical observation. Both of these avenues have seen revolutions in the application of machine learning and AI in recent years. AI systems can formulate axiomatic extensions to existing laws, and the wealth of data available from experiments allows for science to take place in the data.

We are already at the point at which AI systems can infer such things as conservation properties (such as the conservation of energy and momentum) and propose underlying 'laws', given only data. Furthermore, they can propose experiments to gather maximal knowledge from new data. Couple to this logical reasoning capability and the ability to operate at scales well beyond human, and one has a recipe for a genuine automated scientist.

In the coming decade we are likely to see a growth in quantum computation for machine learning. This promises the ability to solve the hardest problems in machine learning and beyond using some of the most bizarre physics we know – which will be transformational.

This is truly the age of the algorithm – and these algorithms are machine learning.



Stephen Roberts
RAEng/Man Group
Professor of Machine
Learning, Department
of Engineering Science,
and Director of the
Oxford-Man Institute of
Quantitative Finance

Making healthcare smarter

By Chris McIntyre

Every solution to a problem seems to create new complications. Improved access to healthcare, rising living standards and technological advances have greatly improved our lifespans and contributed to higher populations than ever before. But rising demand for access to healthcare, coupled with more people living into old age, also places an increasing strain on healthcare providers.

As society becomes ever more mobile, it is important to make sure that patients' records can follow them to make GPs and hospital doctors aware of their full medical history. But moving patient data from paper to computer is yielding far greater benefits than just improved administration.

Researchers led by Professor Paul Leeson, at the University of Oxford's Radcliffe Department of Medicine, have been using machine learning – a form of artificial intelligence – to examine echocardiograms of patients visiting hospital suffering with chest pain. The new system can detect 80,000 subtle changes that would be otherwise invisible to the naked eye, improving diagnosis accuracy to 90% and potentially saving the NHS millions of pounds in avoidable operations and treatment. This is just one of many new applications of AI to healthcare.

'Digital technology is now part of everyday life,' says David Clifton, Associate Professor in Oxford's Department of Engineering Science.

'In healthcare, we have seen an astounding level of hype surrounding the use of AI – but there is real promise for helping people. For example, one thing AI can do better than humans is to assimilate enormous amounts of data, continuously, and use this to spot subtle events in patient data that are otherwise easily overlooked.'

Professor Clifton's team in Oxford is currently working in partnership with several research centres in the UK and in China to create enormous databases of medical data that can be used to develop new generations of complex AI algorithms for healthcare.

'AI algorithms are certainly "data hungry", but the Oxford approach is grounded in ensuring that everything we do is driven by medical

doctors – it is that clinical knowledge, baked into the algorithms, that separates so-called clinical AI from regular AI,' says Professor Clifton.

This approach is being demonstrated at scale by Sensyne Health, a company based on Oxford research formed by Lord Drayson, a former science minister in the UK government. The outputs from the labs of Professor Clifton and Professor Lionel Tarassenko are being delivered into the NHS via Sensyne Health.

The current applications of AI aren't, however, limited to improving diagnosis. Many medical techniques require years of practice to perfect, but some researchers are developing technologies that could enable computers to help to improve the skills of less experienced hospital practitioners.

Alison Noble is the Technikos Professor of Biomedical Engineering in Oxford's Department of Engineering Science. Her main research interest is in biomedical image analysis, with a particular focus on raising the profile of ultrasound imaging, and she has been developing technology to assist ultrasound scanner operators.

'Ultrasound machines are complex devices to master,' says Professor Noble. 'They involve constant interpretation of the data on screen, which directs the actions of the technician performing the scan.'

'Computers don't see data in the same way humans do. While we filter out what we see as noise or static, looking for anything we recognise as a head or a foot, computers can analyse all of the data at once to extract vital clues about what the scanner is actually passing over.'

One of the programs Professor Noble has been developing is able to recognise

the key features that doctors look for in the normal development of babies during routine ultrasound scans of pregnant women. Once the computer has recognised a feature such as the head or a heartbeat, it flags it to the technician who can then move on to look for the next feature.

'This active assistance from the computer is particularly useful for less experienced practitioners, ameliorating the effects of lower levels of training in remote areas where women may not have easy access to hospitals,' adds Professor Noble. 'This technology can be used with a small, portable computer and a handheld scanner, effectively providing patients in remote, rural parts of the world with access to much more accurate healthcare diagnostics than before.'

'It also reduces the need for repeated scans of the same area, making the process safer for the baby.'

The ability of AI to look through the noise in medical scans is also yielding interesting results in preventative healthcare.

Charalambos Antoniades is Professor of Cardiovascular Medicine at Oxford and leads the Oxford Translational Cardiovascular Research Group. His team has developed new technology that analyses coronary computed tomography (CT) angiograms and can flag patients who are at risk of deadly heart attacks years before they occur.

'The standard software currently used with CT scanners is designed to filter out certain types of tissues, such as fat, to make it easier to see organs like the heart,' he says. 'However, huge amounts of data are obtained with each CT scan, which are currently not used because we don't know what they mean. This is what our

This active assistance from the computer is particularly useful for less experienced practitioners, ameliorating the effects of lower levels of training in remote areas where women may not have easy access to hospitals.

research brings to the surface and analyses.'

Heart attacks are usually caused by inflamed plaques in the coronary artery causing an abrupt blockage of blood getting to the heart. Professor Antoniades' team has developed a technology, called the fat attenuation index (FAI), which detects the inflamed plaques prone to causing heart attacks by analysing CT images of the fat surrounding the arteries – something that is filtered out by any standard CT image analysis software.

'This new technology may prove transformative for primary and secondary prevention,' he adds. 'For the first time we have a set of biomarkers, derived from a routine test that is already used in everyday clinical practice, that measures what we call the "residual cardiovascular risk", currently missed by all risk scores and non-invasive tests.

'Knowing who is at increased risk for a heart attack could allow us to intervene early enough to prevent it. I expect these biomarkers to become an essential part of standard CT coronary angiography reporting in the coming years.'

In common with the other research teams that are beginning to employ machine learning and AI in healthcare applications, Professor Antoniades notes that the more data that we can gather from patients now, the better the ability of the FAI technology to predict heart attacks will be in the future.

'The key to improving the diagnostic ability of these technologies is to include data from multiple cohorts in multiple countries,' he says. 'The more data you can put in, and the wider the pool it's collected from, the better the computer will be at discerning what is and what isn't a sign of future health risk.'

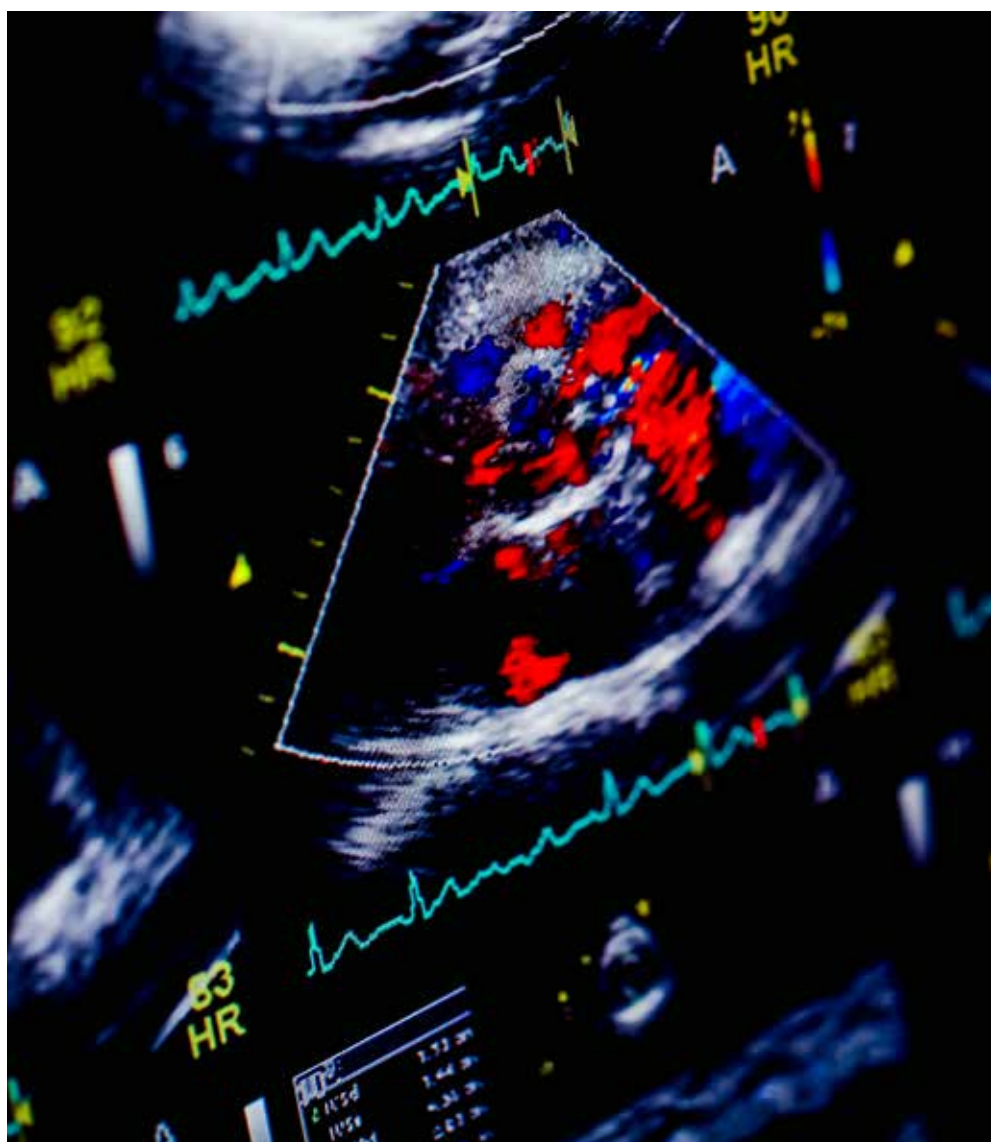


Photo: shutterstock.com

A game changer for drug discovery

By Charlotte Deane

A commonly heard phrase in the pharmaceutical industry is that ‘drug discovery is in crisis’. Over many years the costs for drug development have escalated while levels of success have sunk lower and lower.

Recent estimates suggest that it costs in excess of \$2.5 billion to develop a new drug. This is not just bad news for the profitability of pharmaceutical companies, it is bad news for all of us as it limits the treatments that are available and pushes up the costs of treatments that do exist.

In drug discovery, as in many other areas, AI has the potential to change the game – to make drug discovery quicker, cheaper and more effective, both reducing the cost of development and aiding in the identification of novel medicines.

Drug discovery is a complex multistep process, but it can broadly be grouped into three areas: the identification of targets (these are the naturally occurring cellular or molecular structures involved in the disease); the development of a specific drug molecule that will modulate the activity of that target; and ensuring the end product is safe for humans to take.

AI has been used for decades within computational approaches to drug discovery but has only recently started to offer the types of impacts that could really change the drug discovery pipeline. It is in the area of developing potential drug molecules that we currently have least traction but perhaps most promise for change.

One of the biggest challenges in using AI in this area is the data – both the amount and its heterogeneity and quality. It is difficult and challenging even to obtain data for most of the steps in the drug development pipeline. Using AI in drug discovery is often like training an algorithm to recognise pictures of cats when you have no pictures of cats but a relatively small number of out-of-focus, badly annotated pictures of dogs and elephants.

One way around the data challenge is to use AI techniques on relatively small amounts of high-quality data that are specific to a given target. In standard drug discovery, once a potential drug molecule has been

found, human experts look at all the data available and suggest new candidate molecules that should be more effective or safer. This is an iterative process until the molecules are considered ready for trials. Recent work has shown that an AI algorithm is able to make better candidate suggestions than human experts and so turn a potential drug molecule into a safe and effective version more quickly and more cheaply.

The more general problem is with novel targets and molecules. Where we do not yet have extensive experimental data, this is more challenging for humans and for AI. Could AI predict an effective, safe drug candidate without needing extensive experimentation?

In this context, people have focused on specific tasks within the pipeline – for example, using AI to search the space of potential drug molecules. This is vast – estimated at around 10^{60} (to give an idea of scale, there are only 10^{24} stars in the universe). It is impossible to calculate the properties of all these molecules, but AI is starting to be able to explore this space in a way humans and other types of algorithms cannot. Other types of AI algorithms borrowed from image processing have been used to predict far more accurately than ever before how well a potential drug molecule will bind to a given target, both with and without information on the target.

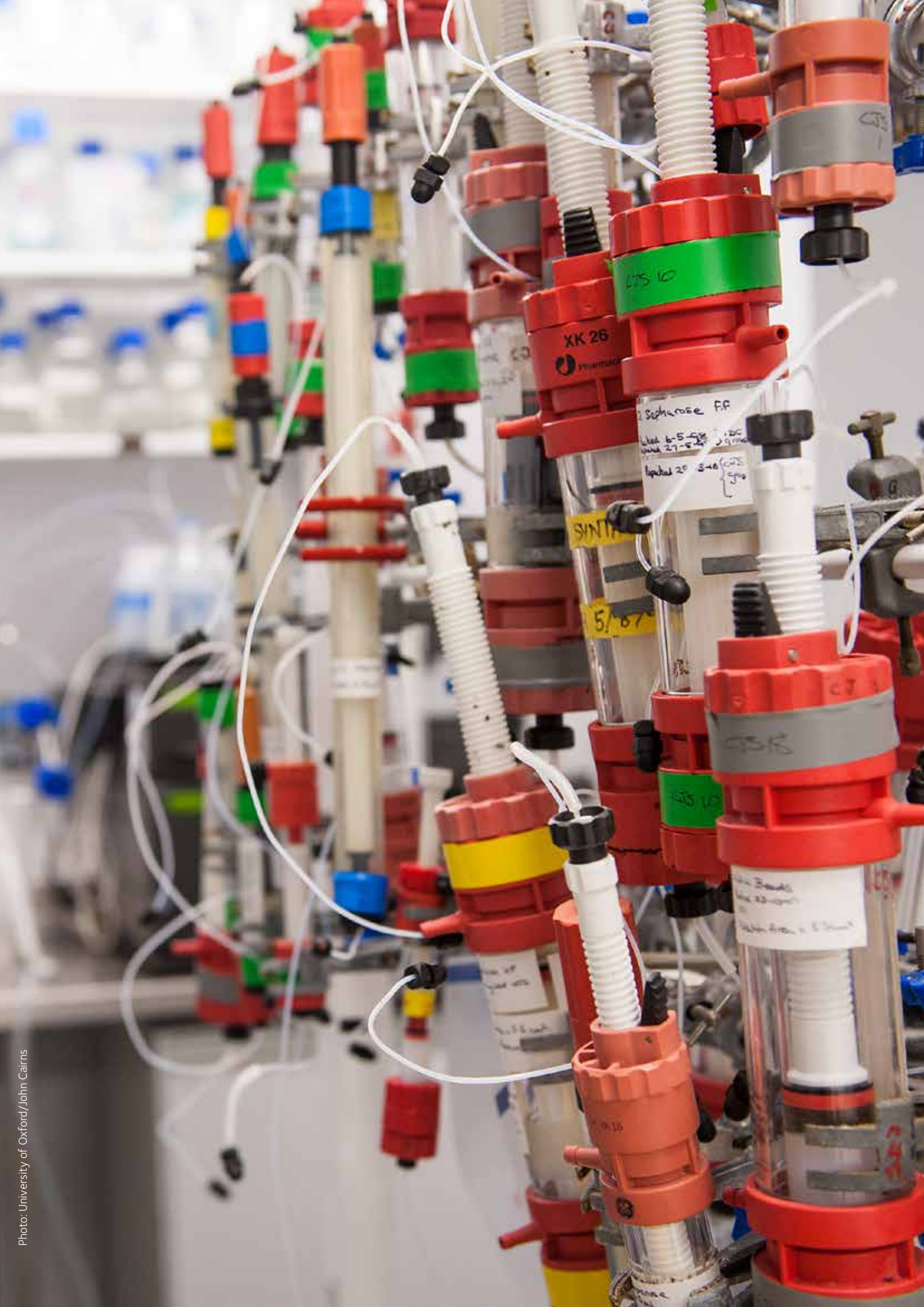
Many challenges remain: none of these methods are accurate to a level that can be used without significant amounts of wet lab experimentation. All of them require human interpretation, and there are still real questions about the generality any of them can or will achieve.

But AI algorithms and techniques are already changing the way drug discovery is done, and as the algorithms improve, as we gain a better understanding of how to handle and represent the data, and also what data to collect, their benefits can only continue to grow.



Charlotte Deane
Professor of Structural Bioinformatics and Head of the Department of Statistics

Using AI in drug discovery is often like training an algorithm to recognise pictures of cats when you have no pictures of cats but a relatively small number of out-of-focus, badly annotated pictures of dogs and elephants.



Using machines to generate biomedical intelligence

By Gil McVean

It's been a busy period here at the Big Data Institute, or BDI, Oxford's recent addition to the rapidly growing biomedical research campus.

Over just a few days we hosted a board meeting for a major pharmaceutical company, held a kick-off meeting for a collaboration with another, gave a tour to representatives from HM Treasury and BEIS, participated in a NICE expert working group on real-world evidence, and were part of a successful Oxford-led bid to establish a hub for AI in biomedical imaging.

It seems that everyone wants to know about AI, machine learning and big data in health research. And it's not surprising. The dramatic advances we've seen in the ability of algorithms to identify and use complex patterns in images, documents, streams of financial data and other data-rich domains are beginning to transform the way in which biomedical and health data-related research can be carried out.

From solving mundane but critical tasks, such as maximising the efficiency of healthcare delivery, to the holy grails of automated drug design or individualised therapy, AI is being deployed across the world with enthusiasm, hype and occasional success.

Within Oxford, we've been fortunate enough to have in place many of the pieces we need to make real the promise of biomedical AI. This includes an incredible history of population health research, leading back to Richard Doll and the British doctors' study on smoking, with an emphasis on clinical trials and population-scale longitudinal measurement; huge strength in the statistical underpinnings of AI, often referred to as machine learning; a community of clinician-scientists who have the insight and drive to understand the need and to help facilitate and shape data-driven research programmes; and a university's worth of fantastic engineers, informaticians, epidemiologists, genomicists and so on, excited by collaborative research and hungry to see their insights make a difference to patients.

The BDI acts as a hub for such activity, supporting the necessary training, computational infrastructure and information exchange, while also leading research programmes ranging from mapping the burden

of antimicrobial resistance across the world, to developing mobile apps for measuring the parts of memory that are fastest to decline in dementia.

To a large extent the needs of an AI-driven research programme in healthcare are not so different from any other data-driven problem. Take, for example, the challenge of automated feature prioritisation from imaging modalities, such as pathology or radiology. We want the computer to help the clinician spot features of importance, building on sets of expert-curated training data, coupled with learning algorithms that improve with experience. This requires a close loop between the engineers, clinicians, algorithm developers and, of course, access to the critical high-quality data sources.

This is the type of problem AI has proved hugely competent at solving – it's a game, like chess or Go, where the rules are set and the machine has to learn the best strategies. Clearly, issues such as repeatability, reproducibility and generalisability are important, but we don't necessarily require the machine to explain why a particular decision has been made. We just need a good decision, fast.

But many of the core problems in biomedicine are fundamentally different from this class of task. Consider the problem of investigating whether some patients respond better to one type of drug than another. Resources, such as the UK Biobank, which are measuring vast amounts of biological, clinical, behavioural and medical data on hundreds of thousands of people, give unprecedented power to find complex patterns. So if we were to use AI to ask whether there are differences in the medical trajectories between those patients given drug A or drug B, the answer would almost certainly be yes. Put another way, by looking at the entirety of a person's data, I can probably work out whether they were given drug A or drug B with reasonable confidence.

But that doesn't necessarily mean that these differences were the result of taking the different drugs. Perhaps drug A is more often given to those who are likely to do well because they

have fewer other diseases, or because its use just happens to be preferred in a couple of hospitals that have particularly good specialists and care pathways for the disease.

In this and many other medical problems, the critical intelligence we need is an understanding of causality – the health benefit likely to arise from a particular intervention. And this doesn't fall naturally from AI. Rather, it is something that only clinical trials, despite their cost and time, can assess.

So what is the role of AI in such work? There are two key areas, both of which the BDI is pursuing. First, we can use AI to make us much smarter about generating therapeutic hypotheses to take to trials, building on a growing wealth of data types that give us clues to causality (such as genomics, longitudinal data, experimental screens and high-resolution biological measurement). Second, we use AI to make trials themselves better, by finding the patients most likely to benefit, the readouts able to measure impact the fastest, and by analysing the clinical data arising to refine hypotheses and iterate.

AI is ultimately just a tool, but it's one that allows us to do science better and get the benefits out into the real world faster.



Gil McVean
Professor of Statistical Genetics,
Nuffield Department of Medicine,
and Director of the Big Data Institute

The Oxford spinout company using AI to diagnose heart disease

By Stuart Gillespie



Save lives, and save money for the health service. It's an outcome few could complain about, and it's happening because of artificial intelligence.

The medical diagnostics company behind this development, Ultronics, was spun out from Oxford University research in 2017. Using the power of AI, Ultronics aims to improve the accuracy of echocardiogram interpretation to above 90% – substantially better than the 80% currently achieved by human doctors.

This, say the company's founders, will save lives by identifying more people at risk of heart disease and – by reducing the number of patients unnecessarily sent to theatre – potentially save billions for health services around the world.

Paul Leeson is Professor of Cardiovascular Medicine in Oxford's Radcliffe Department of Medicine and one of the founders of Ultronics. He says: 'Echocardiography is the most widely used imaging test in people with heart disease. In most hospitals, over ten times more echocardiograms are performed than any other imaging test in cardiology. This is because echocardiograms can be performed quickly, anywhere in the hospital, including at the bedside or in the clinic. Echocardiograms are also performed in the community and in remote locations, or areas where resources are limited.'

'However, you need an expert to interpret the images and reach a diagnosis. When the expert is good, then the test can be very accurate. But because levels of experience vary, this can be difficult to control.'

'We wanted to fix this by using AI methods to standardise how images are analysed, lifting the quality of interpretation so that it is always as good, or better, than an expert reader. To do this, we built up databases of hundreds of thousands of echocardiography images linked to information about what was unique about the person who was being imaged and what happened to them over time. By combining machine learning with clinical know-how, we were able to identify associations between features hidden within the echocardiography images and what happens to patients. Doctors can then use this information to decide how to look after the patient.'

Ultronics' co-founder and CEO Ross Upton is, perhaps unusually in a University

spinout company, a current graduate student at Oxford, nearing completion of his DPhil in cardiovascular medicine under Professor Leeson. Upton had the idea of applying AI and machine learning techniques to this field after learning of the shortfall in the accuracy of diagnosis. Within two years, Ultronics had been spun out of the University with the help of Oxford University Innovation – Oxford's research commercialisation arm – attracting more than £10 million in investment led by the Oxford Sciences Innovation fund.

Upton says: 'The first product of Ultronics, EchoGo, is based on extracting features from stress echo images and using a supervised machine learning model to predict the outcome of a patient one year following the test. The features we extract from the images are all biologically relevant to the disease process – some of which are clinically known and others which are entirely novel features that we have patented.'

'We used one-year patient outcomes as the gold standard, rather than how someone has reported the scan, because we know operators interpret the scan correctly only 80% of the time. We therefore need to follow up the research participants for a year after the exam to see what actually happened to them after the test. If the test is interpreted incorrectly, the patient would get sent for an angiogram unnecessarily; if the test was reported as normal but there was underlying disease, then the patient would get sent home when they should have been sent for an angiogram. It's these errors that EchoGo is going to reduce.'

Professor Leeson adds: 'Stress echocardiography is used widely across the world – it is the most commonly used functional imaging test for coronary artery disease in the UK. By using the AI technology to ensure consistent and accurate interpretation, you can reduce the need for unnecessary additional investigations and ensure you do not miss disease. This improves the care of the patient and significantly reduces costs for the NHS. Also, because stress echocardiography uses ultrasound equipment that is already available in hospitals and can be delivered by

existing clinical staff, it means hospitals can more carefully consider whether they need to spend money on expensive new tests and infrastructure or instead put their existing infrastructure to better use.'

The next step for the company, says Upton, is to achieve a CE mark and clearance from the US Food and Drug Administration so that EchoGo can be introduced to clinics and begin improving patient outcomes. He adds: 'We are also looking to expand our already large-scale clinical trial to 30 different hospitals across the NHS. The next innovation is to completely automate EchoGo, which will help provide an instantaneous result to clinicians. This will be done by utilising newer deep learning frameworks, which are being refined at the moment by our research and development team. Following that, we will look to tackle other disease areas within echocardiography, such as heart failure and valve disease.'

Reflecting on the process of spinning out a commercial company from University research, Professor Leeson says: 'A lot of companies are spun out from Oxford, but that is not because it is an easy thing to do. The number reflects the amount of high-quality, truly "translatable" research being carried out by investigators in departments. This is coupled with very effective and experienced support from Oxford University Innovation. From concept to spinout took us two years, and we had to get over a range of hurdles on the way, including securing IP and patents, being awarded pre-spinout seed funding to build aspects of the technology that would be attractive to investors, and, finally, convincing a lead investor to invest in both the technology and us, as founders.'

'You have to have a really game-changing idea, with science to back it up, to convince investors. Even at that stage, negotiating the details of how the company is formed and its ongoing relationship with the University can take several months to arrange. After that, the Oxford environment, with supportive backers such as Oxford Sciences Innovation, means the acceleration and growth of the company can be very rapid.'

Self Driving



Training autonomous vehicles using real-life human behaviour

By Stuart Gillespie

Driverless cars are on their way – there’s little doubt about that. But before they hit the UK’s roads, they need to be tested in realistic simulations to ensure that this transformative technology will be a safe and positive addition to our lives.

Latent Logic, an Oxford University spinout company, is helping enable this vital testing with technology that teaches autonomous systems using real-life examples of natural human behaviour on the roads.

Professor Shimon Whiteson of Oxford’s Department of Computer Science, co-founder and chief scientist at Latent Logic, explains: ‘Autonomous vehicles must be tested in simulation before they can be deployed on real roads. To make these simulations realistic, it’s not enough to simulate the road environment; we need to simulate the other road users too: the human drivers, cyclists and pedestrians with which an autonomous vehicle may interact.

‘Latent Logic is using a machine learning technique called imitation learning to build realistic human behaviour models. These models make it possible to test autonomous vehicles quickly and robustly.’

Latent Logic grew out of an EU-funded research project that trained semi-autonomous telepresence robots to behave in a socially ‘normal’ way. Since it’s difficult to quantify what is meant by socially normal, it’s much easier to train such systems to imitate the behaviour of humans acting in a socially normal way.

Professor Whiteson decided to explore the commercial potential of this technology, recruiting postdoctoral researcher Dr João Messias to be co-founder and chief technology officer. Kirsty Lloyd-Jukes then joined the company as CEO.

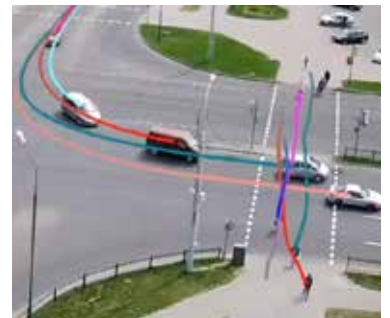
The technology works by combining state-of-the-art computer vision with imitation learning. Professor Whiteson says: ‘Our models extract the “latent logic” behind real-life examples of natural human behaviour. As a result, they can respond realistically even in new situations.

‘We use computer vision to collect these examples from video data provided by traffic cameras and drone footage. We can detect road users, track their motion, and infer their three-dimensional position in the real world. Then, we learn to generate realistic trajectories that imitate this real-life behaviour.’

By providing a service that enables better training and testing of autonomous vehicles, Latent Logic hopes it can hasten the safe introduction of what will be life-changing technology. Professor Whiteson adds: ‘While many of the biggest players in this market are international, there is also a lot of energy in the UK in this sector, and Oxford is a hotbed of talent and entrepreneurship in machine learning, robotics and autonomous vehicles.

‘Autonomous vehicles are improving rapidly, but we are still some way from realising the dream. It is not just about perfecting existing technology – there remain fundamental unsolved problems in building sufficiently robust autonomous systems. At Latent Logic, we believe our technology can play a critical role in addressing those unsolved problems.

‘In addition, our technology has numerous other applications. Situations in which you want socially normative behaviour are great candidates for learning from demonstration, as is robotics – from factories to warehouses to homes. You can also think about video games, where people might want to play against bots that can imitate the style of their favourite professional gamers. The sky’s the limit.’



Latent Logic grew out of a research project which trained semi-autonomous telepresence robots to behave in a socially ‘normal’ way.

Transforming the finance sector

By Chris McIntyre

The effects of technology on the way we work have been widely discussed in recent years, particularly as advances in smart technology become apparent in our daily lives.

AI is already playing a role in the finance sector, from fraud detection to algorithmic trading and customer service, and many within the industry believe this role will develop rapidly within the next few years.

Hedge fund managers were among the earliest adopters of machine learning several decades ago, recognising the advantages of using technology to analyse market data, both in terms of speed and volume.

Professor Stephen Roberts is Director of the Oxford-Man Institute of Quantitative Finance and Professor of Machine Learning in the University of Oxford's Department of Engineering Science, and an expert on the application of machine learning approaches to data analysis in finance.

'Most of the techniques that form the backbone of modern AI and machine learning, that really have impact in industry and commerce, have their roots in algorithms that were known about 20 years ago,' says Professor Roberts. 'What has changed dramatically in the last two decades is computing power and data volume.'

'Traders today have access to a reservoir of data that is beyond human ability to analyse without computers – everything from real-time shipping data and weather reports to global commodity demand and regional news updates.'

Traders use algorithms to distil insight from this data, which, combined with their own understanding of financial markets, can help them make better decisions, hedge risk and make the right calls in terms of the assets that they are trading.

'Algorithms are extremely good at teasing out the patterns and correlations in these very large, unstructured, disparate datasets, then proposing potential trade options to people who can implement them live into the markets,' adds Professor Roberts. 'This is a discovery that I think we've only seen the tip of the iceberg of.'

Despite the superior analytical power of AI systems, Professor Roberts doesn't see humans being completely replaced any time soon – but he does see the way we work continuing to be reshaped by technology.

AI is augmenting human capability, acting

as an extra conduit of knowledge and helping professionals make the right decisions in a quicker time, and this aspect of AI is a big area of development at the moment.

'It is also opening up new products and trade opportunities. Finance houses are extending their remit of data acquisition to include the weird, wild and wonderful information, not just price series that would have been the mainstay of finance a few decades ago. This could include possible links between social media sentiment in a country and the output of products from that country.'

Nir Vulkan is Associate Professor of Business Economics at Oxford University's Saïd Business School, as well as the creator and director of the Oxford Online Programme on Algorithmic Trading and the FinTech programme. He also sees AI as increasingly augmenting the skills of traders rather than replacing them.

'Algorithmic trading is a powerful tool for traders with experience and good knowledge of their sector, but it can be more risky if used by less experienced traders,' he says.

'A major advantage is that algorithmic trading removes the emotion from trades, helping to guide traders who may be nervous or excited. A computer only looks at data, and most successful trades are based on solid market data.'

'Over their careers most traders develop a set of rules that guide them. AI is a natural development of this method, which is why most traders are happy to use AI tools.'

Like Professor Roberts, he has also seen the application of AI opening up new opportunities across the financial sector.

'The financial sector is going through a "syntax revolution" at the current time,' says Professor Vulkan. 'Pressure on regulators from the government to work with startups rather than against them has led to an explosion in their numbers. New companies such as Funding Circle, Monzo and TransferWise are offering new ways of funding that weren't available previously.'

'All of these companies are based on algorithmics, enabling them to use technology to remove barriers to funding through the same

application of AI to lending as traders have been using for investments for years. London is leading the way in this sector.'

Following this trend, Professor Vulkan notes that banks are also starting to use this technology in the services they provide to customers. Banks have been buying up the most successful of these smaller companies to incorporate their technologies into their own service offering. For example, HSBC has introduced an app that nudges customers if they have spent more or less than usual based on previous spending habits, to alert them to any problems early.

'This is a particularly exciting area of growth because this sector has needed more funding for a long time,' he adds. 'Banks have not been innovative enough or lent enough for some time now.'

This technology helps banks to reduce their exposure to risk while also making funding available to people and businesses that were excluded from lending before. AI is making the lending market both more effective and inclusive, encouraging new businesses to grow. It has become apparent that lending to risky ventures works, which is good news for entrepreneurial students.

'Britain is ahead of the curve in this sector,' says Professor Vulkan, 'and it's rewarding to see the role that Oxford spinout companies have played in making that happen.'

A major advantage is that algorithmic trading removes the emotion from trades, helping to guide traders who may be nervous or excited.



Saving the planet

By Lanisha Butterfield

A combination of ever-increasing population growth – 7.8 billion and counting – and short-sighted human behaviours have left the natural world at tipping point. As a result, multiple species now face extinction.

As scientists attempt to prevent what could be a devastating biodiversity crisis, they are looking towards a surprising tool: artificial intelligence.

Significant strides in technology and AI development have enabled new research opportunities in areas of conservation and ecology, such as exploring extreme environments like the deep sea, deserts and the poles in order better to understand and protect the species that live there.

At Oxford University alone, multi-faceted approaches to using AI have allowed academics from across the institution to take their research to the next level: projects as diverse as the conservation of penguin populations in Antarctica and wild lions in Africa, helping governments crack down on poaching and the illegal wildlife trade, and the development of algorithms that will help scientists identify and track wildlife species from their seismic vibrations.

Although very different issues at first glance, these projects share a common thread: an interdisciplinary approach which allows the research to tackle real-world issues.

Conservation isn't just about saving specific species; it's about understanding the factors involved in human-animal conflict and how these issues impact people's everyday lives. This needs to be approached on two levels, beginning by getting to grips with the ecosystem around an issue, its importance to local communities – be that cultural, spiritual or economic – and the demands placed on it by different stakeholders. The next step is to understand how best to work with the people most affected by a conservation decision, as well as other groups and experts, to protect both biodiversity and people's wellbeing.

'A top-down approach can be really counterproductive in tribal communities. It's important not to judge and instead see if there is anything that you can do to make the wildlife more of a social benefit than a pest,' says Dr Amy Dickman of Oxford's Wildlife Conservation Research Unit (WildCRU), founder of the Ruaha Carnivore Project, which works to protect the lion prides of southern Tanzania by offering incentives to the community. Since the project launched six years ago, carnivore attacks on livestock in these communities have been reduced by 60%,

and big cat killings have decreased by 80%.

As daily life and digital life have become increasingly connected, conservation research has come to include working to understand both offline and online footprints. In the case of the Oxford Martin Programme on the Illegal Wildlife Trade, which aims to tackle the trade of endangered species, this work involves understanding how organisers use the dark web as a sales forum.

The illegal wildlife trade poses a severe and ever-growing threat to global biodiversity, responsible for a 60% decline in elephant numbers alone between 2009 and 2014. It is also a big money earner, generating up to \$10 billion a year for those behind it.

The Oxford Martin Programme brings together experts in areas ranging from international development and economics to computer science, psychology and engineering. This multi-disciplinary expertise enables the development of tools for the surveillance and tracking of online wildlife sales, as well as the unravelling of the motivations driving the trade.

The programme's Dr Joss Wright explains: 'Our project involves colleagues from across the University whom you wouldn't naturally expect to work together. But each area of expertise is a necessary puzzle piece, essential if we are to see the bigger picture and have tangible societal impact.'

AI-powered technology is fundamental to the project's interdisciplinary approach, allowing researchers to look at the illegal wildlife trade with fresh eyes and understand why so many communities support and depend on it. From these observations they can work out incentives to help communities protect species, by making the wildlife as valuable to local people alive as it has become dead. Researchers from the Oxford Internet Institute have developed algorithms that allow them to understand how the web enables this illicit trade. Then, working in collaboration with practitioners, they implement their findings with a view to changing policy in ways that support the needs of communities.

Of the role that AI has played in the project and how beneficial it has been in achieving the stated goals, Dr Wright says: 'For a long time this work was done by people manually, over a much

longer period of time. But technology and AI developments mean that computers can now do some of this work for us: in our case, trawling the depths of the internet for patterns and shifts in how people engage with the illegal wildlife trade. This includes, for example, the kinds of forums used for trade, and changes in terminology which could potentially be linked to breakthroughs in law enforcement and then having to adapt the language used in order to go undetected.

'These developments have allowed us to make more detailed and complex predictions, and to use huge amounts of data that we would not have been able to five years ago.'

AI does not just help scientists to understand a conflict or research an area through data analysis, it also enables smart technologies that allow them to monitor animal behaviours. By training computer algorithms on a set of available data, machines can now learn what they should do for a given challenge – such as classifying photographs by the species found in them, identifying areas of a satellite image containing water or intact forest, or translating speech from one language to another.

An interdisciplinary approach combined with AI-powered technologies has been key to the success of the Ruaha Carnivore Project. The project has initiated a number of interventions that support the protection both of lions and people locally.

Dr Dickman says: 'Lions have incredible international value, but in most places they are currently worth more to local people dead than they are alive. If we want locals to protect their wildlife, they have to value it as something that actually brings them some benefit – and that means giving it a tangible value. If the presence of this wildlife improves their lives, they are going to want to keep it there.'

In Ruaha, killing a carnivore earns people – particularly Barabagi and Masaai tribesmen (warriors) – respect, status and even gifts like cattle from the rest of the community. So the team has worked to build relationships with local people and to understand what it would take to make them stop this behaviour. Their concerns, particularly among the women in the group, were the same as they would be for any of us: economic stability, healthcare, education for their

The Zooniverse platform now hosts more than 100 diverse projects, with one of the most popular being the ecology initiative Penguin Watch.



Photo: pexels.com

children and veterinary health for their livestock.

The team developed programmes in all of these areas, including the opportunity to become a 'lion defender': employing community warriors to defend rather than kill lions, and offering them a monthly wage in return. As part of this work the team introduced community camera trapping, where the lion defenders run and monitor automated camera traps (cameras powered by AI which automatically take photos when an animal passes), and the villagers receive points for every image captured. These points translate into extra money for the community – around \$5,000 per village – every three months, which goes towards healthcare, education and veterinary medicine. This programme has demonstrated that the presence of wildlife on village land can be an asset that directly generates important community benefits.

Automated camera traps also drive another medium that has become an essential aid to conservation and ecology research: citizen science.

Founded by Oxford University academics ten years ago, the Zooniverse platform runs on support from volunteer 'armchair scientists' who help the team with their research by identifying and classifying everything from images of specific animal species, to galaxies in space and regions affected by hurricanes and extreme weather – all from the comfort of their own homes.

Thanks to the internet and technological developments like the evolution of the smartphone, the model of involving public volunteers in research has progressed

greatly in the past ten years.

Support from the public enables researchers to process data significantly faster. Over time it has become vital to conservation research streams. The Zooniverse platform now hosts more than 100 diverse projects, with one of the most popular being the ecology initiative Penguin Watch. The programme uses a series of time-lapse cameras, set up across the southern hemisphere, to capture data on penguin behaviours and breeding habits.

Oxford penguinologist Dr Tom Hart, from the Department of Zoology, says: 'Our cameras are already taking our place in the field, and are in fact better than us at the legwork. They take hourly pictures without fail, and all we need to do now is make them a bit smarter and more interactive.'

'We work with researchers in Oxford's Department of Engineering Science to build our own camera technology, using machine learning to program them to operate independently – much like a Mars rover. If the lens were to become covered with snow, we would teach the camera to recognise the white covering and to heat the lens, melting the snow away. Penguin behaviours are also quite synchronous, so if anything unusual happens within view, the camera will detect that something new is happening and take more photos of it.'

However, although machines are now able to carry out a lot of the legwork for scientists, the researchers themselves feel strongly that they could never and should never replace them entirely.

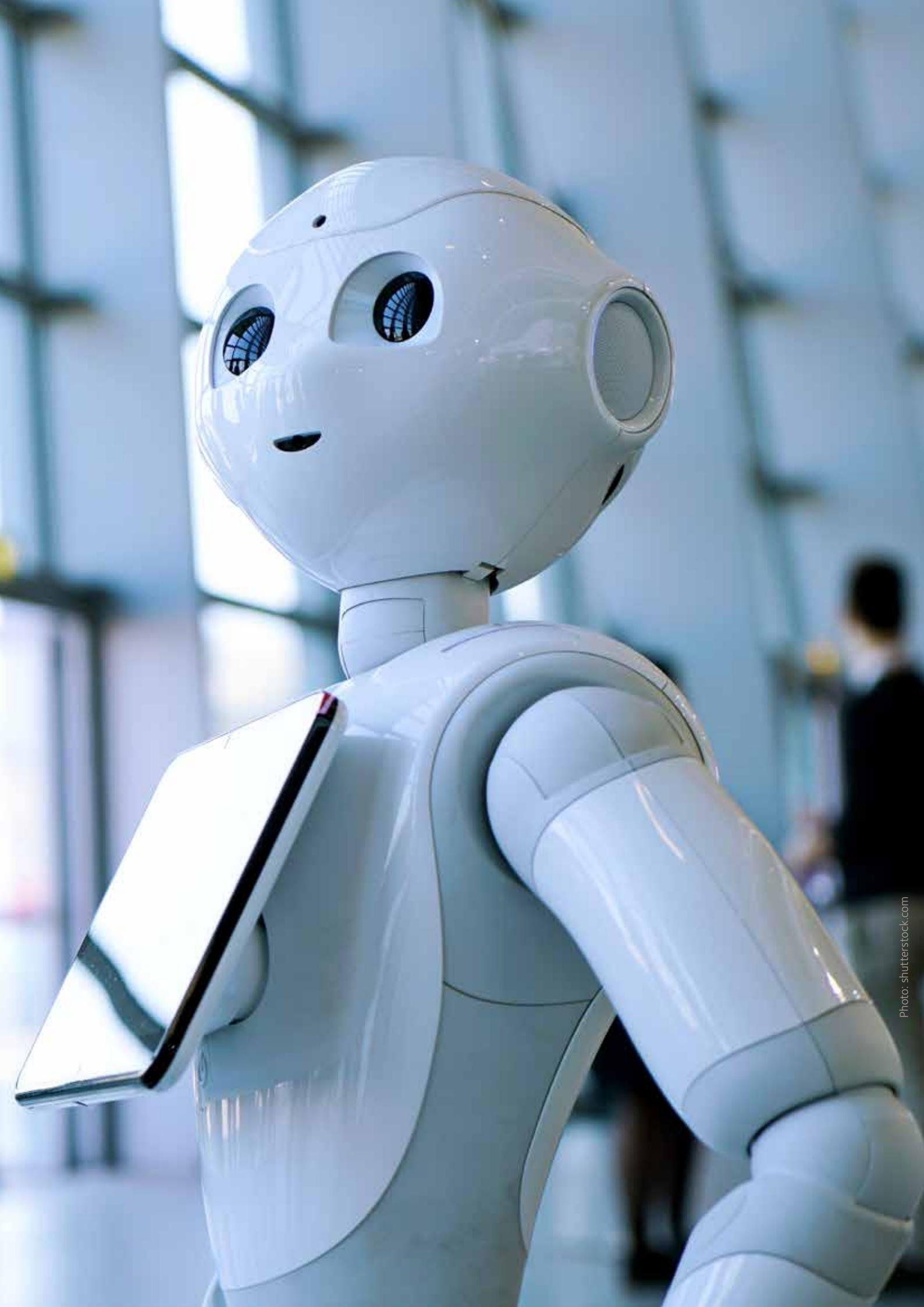
Dr Wright says: 'People automatically assume that they can trust computers to do intelligent work. But algorithms with built-in discriminatory bias that can affect people's lives – particularly minorities and vulnerable groups – are a big concern that affects conservation work as much as it does other areas.'

One potential way of reducing the online illegal wildlife trade is to use automated filters to block content, such as adverts, linked to the given subject matter. These filters can, however, have discriminatory biases built in that can negatively affect disadvantaged groups. An algorithm trained to block illegal wildlife trade postings in Kenya, for example, could be much more likely to block innocent adverts for animal-related products in the country. An embedded mistake like this, learned from poor data used to train the algorithm, could significantly harm legitimate sellers on the basis of their location alone.

'It is a mistake to trust computers too much, or to knowingly let them make decisions that will negatively affect people's lives,' says Dr Wright. 'You want to avoid building systems that exacerbate bias by design.'

He adds: 'You always want to keep a human in the loop, so that they can pick up when a computer makes a mistake, which it inevitably will.'

'Our work would be very different without AI, and it would definitely take much longer. But we could never rely on it entirely. People have an understanding that AI could never provide – we need to use the two together.'



Robots thinking fast and slow: a case for embodied artificial intelligence

By Ingmar Posner

These days, artificial intelligence is everywhere. We routinely interact with voice assistants. AI technology increasingly out-diagnoses our most experienced doctors. And autonomous cars have been just around the corner for almost a decade. But where, you may ask, are the robots?

Where are those machines that work for, with and alongside us? Why can I buy a voice assistant but not a robust and versatile household robot?

The answer lies in the fact that embodiment – the notion of a physical agent acting and interacting in the real world – poses a particular set of challenges.

In the Oxford Robotics Institute we address the full gamut of challenges facing embodied agents, from autonomous driving and long-term autonomy in complex, dynamic environments via the deployment of manipulators and legged vehicles, all the way through to the development of sophisticated sensing technologies like robot skin.

The Applied AI Lab in particular investigates core challenges in robot learning for the real world. How can a robot perform complex tasks in real-time? How does it know when it doesn't know? How can it efficiently and robustly acquire new knowledge?

Recent advances in AI have built significant excitement as to what our robots may be able to do for us in the future. Machines are now able to play the Atari suite of games or even the massively complex game of Go. Progress is truly inspirational. However, success here relies on the ability to learn cheaply, often within the confines of a virtual environment, by trial and error over as many episodes as required.

This presents a significant challenge for embodied systems acting and interacting in the real world. Not only is there a cost (either monetary or in terms of execution time) associated with a particular trial, thus limiting the amount of training data obtainable, but there also exist safety constraints which make an exploration of the state space simply unrealistic: teaching a real robot to cross a real road via trial and error seems a far-fetched goal. What's more, embodied intelligence requires tight integration of perception, planning and control. The critical

inter-dependence of these systems, coupled with limited hardware, often leads to fragile performance and slow execution times.

In contrast, we require our robots to operate robustly in real-time, to learn from a limited amount of data, to make mission- and sometimes safety-critical decisions, and occasionally even to display a knack for creative problem solving.

Psychology and cognitive science suggest that there are a number of mechanisms in humans that allow us successfully to act and interact in the real world. One prominent example is dual process theory, popularised by Daniel Kahneman's book *Thinking Fast and Slow*. Dual process theory postulates that human thought arises as a result of two interacting processes: an unconscious, involuntary – intuitive – response and a much more laboured, deliberate reasoning.

In the Applied AI Lab we posit that recent advances in deep learning have – for the first time ever – put a dual process theory for robots firmly within reach. For example, our recent work has shown how we can endow a machine with a notion of physical intuition by letting it observe how the world evolves. We can also ask a machine to capture intuitively how a human drives by capitalising on the many demonstrations routinely provided. Much of this intuition is executed significantly faster than the more laboured oracles that produced the relevant information in the first place.

Inspired by findings from across a variety of fields including machine learning and AI, as well as neuroscience and cognitive science, the Applied AI Lab is now building on this work to explore – as one of its research directions – where a dual process theory for robots might lead.

Real-time performance; robust, safety-critical decision making; creative problem solving; tool use: we believe that robots should indeed be thinking fast and slow.

Our recent work has shown how we can endow a machine with a notion of physical intuition by letting it observe how the world evolves. We can also ask a machine to intuitively capture how a human drives by capitalising on the many demonstrations routinely provided.



Ingmar Posner
Professor of Engineering Science,
Department of Engineering Science,
and Head of the Applied AI Lab,
Founder and Deputy Director, Oxford
Robotics Institute

Automation and the future of work: understanding the numbers

By Carl Benedikt Frey and Michael Osborne

In 2013, we published a paper titled *The Future of Employment: How Susceptible Are Jobs to Computerisation?* which estimated that 47% of jobs in the US are at risk of automation.

Since then other, similar studies have emerged, arriving at different numerical conclusions but built on the same intuition – that the future of work can be inferred by observing what computers are capable of. There are good reasons to believe that this view is correct. Back in 2003, MIT researchers David Autor, Frank Levy and Richard Murnane highlighted the disappearance, since 1980, of jobs that were intensive in ‘routine’ tasks.

Their findings were entirely predictable. As early as 1960, Herbert Simon predicted the decline of routine jobs in his essay ‘The Corporation: Will It Be Managed by Machines?’ He argued that computers held the comparative advantage in routine, rule-based activities that are easy to specify in computer code. Through a series of case studies from the same year, the US Bureau of Labor Statistics arrived at a similar conclusion, suggesting that a little over 80% of employees affected by contemporary technological advances would be in jobs involving filing, computing, machine operations such as tabulating or keypunching, and the posting, checking and maintaining of records.

Our estimates – particularly the 47% figure – have often been taken to imply an employment apocalypse. Yet that is not what we were saying. Our study simply looked at the susceptibility of existing jobs – 702 occupations, comprising 97% of the US workforce – to recent developments in emerging technologies such as artificial intelligence and mobile robotics. It did not predict a timeframe, and it did not explore the new sectors and roles that will undoubtedly arise in the years and decades to come.

Appropriately – or perhaps ironically – most of our analysis was carried out using AI and machine learning techniques. First, though, we gathered a group of machine learning experts to assess, in the context of current technologies, the potential automatability

of 70 occupations using detailed task descriptions. Our trained algorithm was then able to assess the automatability of a much wider range of occupations, using data derived from the vast O*NET online jobs and skills database.

We argued in our subsequent report, for instance, that even many non-routine tasks, such as legal writing or truck driving, will soon be automated. Telemarketing and insurance underwriting were among the occupations deemed at greatest risk of automation; social work and many medical professions among the least. Waiting staff were found to be at high risk – an assertion our expert panel did not necessarily agree with but which was proved correct a few years later with the launch of a completely waiter-less restaurant chain. We also provided concerning evidence that jobs associated with low wages and low educational attainment have a strong relationship with potential computerisation.

What our results show is that the potential scope of automation is vast, just as it was on the eve of the Second Industrial Revolution, before electricity and the internal combustion engine rendered many of the jobs that existed in 1900 redundant. Had our great-grandparents undertaken a similar assessment at the turn of the 20th century, they would probably have arrived at a similar figure. Back in 1900, over 40% of the workforce was employed in agriculture. Now it is less than 2%.

Seen through the lens of the 20th century, our estimate that 47% of jobs are exposed to future automation does not seem extraordinarily high. Policymakers need to understand the thinking behind the numbers in studies like ours to draw their own conclusions about the scale of the changes facing us. The world of work is, once again, changing at pace, and will continue to change. We need to be able to craft appropriate responses.

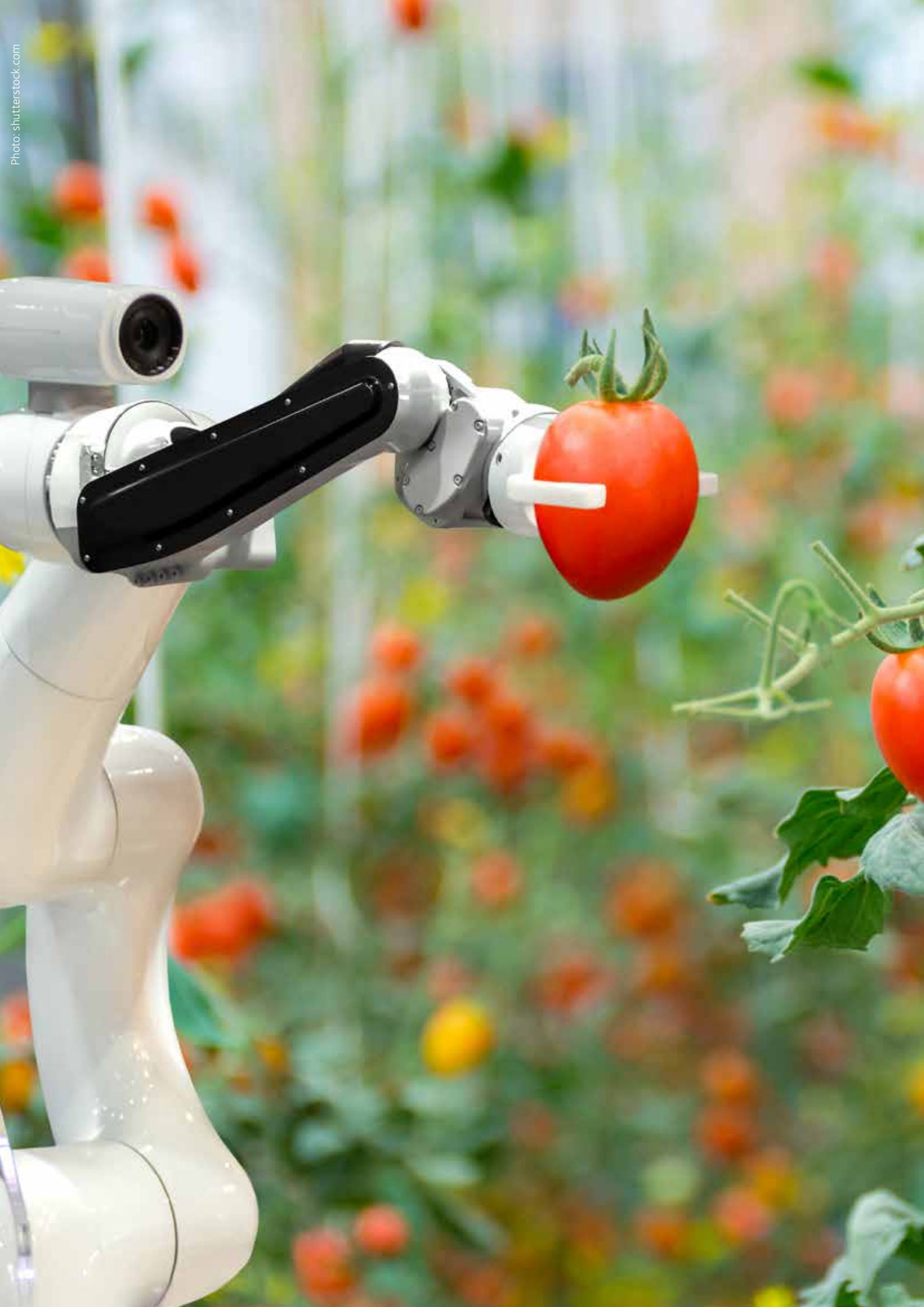


Carl Benedikt Frey
Co-Director and Oxford
Martin Citi Fellow, Oxford
Martin Programme
on Technology and
Employment



Michael Osborne
Dyson Associate Professor
in Machine Learning,
Department of
Engineering Science

Policymakers need to understand the thinking behind the numbers in studies like ours to draw their own conclusions about the scale of the changes facing us.



Female
Age 65



Think about ethics now and AI can empower rather than oppress us

By Sir Nigel Shadbolt

If the coming AI revolution is to be ethical, the right to consent and to understand how our data are being used are fundamental precursors.

It's the Hollywood model that worries people: AI as Terminator robot; mad, bad and dangerous to know.

We talk about the weaponisation of AI, the threat to privacy, the potential loss of jobs.

Those are legitimate concerns, but we should not lose sight of one fact: as in every era of major scientific and technical advance, we have choices about how we use the technology – and we need to make the right choices about AI and intelligent machines.

People often use terms such as general AI, or the digital singularity, to describe a version of AI that is going to take over the planet and threaten humanity's very existence. But we're a long, long way off being able to realise that type of consciousness and self-awareness in machines. The threat right now isn't artificial intelligence – it's natural stupidity.

The chemical industry that blossomed in the late 19th century brought the threat of chemical warfare but also important advances in agricultural fertilisers. Nuclear technology has given us both the atomic bomb and a new source of power. We are now beginning to realise that the advancement of artificial intelligence has the same duality – a duality that we can address now in the systems we're building.

Take healthcare, which AI will transform. The explosion of data relating to our fundamental biological make-up will give AI algorithms the opportunity to decode how the stuff of life is built, how it functions, why it goes wrong and how it might be improved. We'll have an extraordinarily powerful set of tools to help us understand and predict the consequences of everything from our genetic inheritance to our lifestyles.

But how do we ensure that your health data, my health data, are being given and used with consent, and in ways that are appropriate? This is one of the key questions facing us as we stand on the edge of another technological revolution. The worry is that a few very big companies will be in possession both of huge amounts of data and the AI capability to use that data in a way

that may or may not be in the best interests of those who have provided it.

If the coming AI revolution is to be ethical, the right to consent and to understand how our data are being used are fundamental precursors. Privacy is not dead, and the data we generate and which relates to us should not be used without conventions and norms, rules and regulations. We are not powerless to address these and other challenges in the face of impending change.

Responsibility must be engineered into these systems from the outset, and the ethical issues must not be an afterthought. Developers and entrepreneurs, AI scientists and engineers need to think about the ethics of what they do, understanding the consequences of their impressive creativity and its place within relevant legal and regulatory frameworks. Oxford is fortunate to have leading experts in the fundamental technologies underpinning AI, in the societal application of these technologies, and in the ethical and legal issues that surround them. Working together, we can ensure that we get not only the best AI systems, but ethical systems: systems that are proportionate, equitable and transparent. Computer scientists will need to work with other disciplines to ensure this occurs.

It makes sense: medical practitioners have always had to be cognisant of ethics and morality in their profession, and safety concerns come with the territory for aeronautical engineers. What makes us human is not just our ability to apply logic, mathematics and statistics to solve problems – AI is doing this in ever more specific areas, as we saw when a computer program beat an international champion at the board game Go – but our ability to consider these things in a social, moral and ethical context.

Concern about the deployment of AI requires us to reflect on the values we seek to protect in an open society. Values that

in Western thought were framed by the philosophy of the Enlightenment. A preference for reason and evidence, transparency and tolerance, privacy and autonomy, dignity and self-determination. These values are ethical in nature and need constant attention and eternal vigilance if they are to endure. The great Enlightenment moral philosopher Immanuel Kant wrote: 'Act in such a way that you treat humanity, whether in your own person or the person of any other, never simply as a means but always at the same time as an end.' This is a principle that we would do well to apply whenever we deploy the fruits of our technology.

AI should empower, rather than oppress us. Exponential increases in fundamental computing power and available data will give us remarkable abilities to anticipate and model the world around us, but the moral choices about what to do with those insights, what to do with that technology, remain with us as human agents. We cannot delegate those choices to machines.



Sir Nigel Shadbolt
Principal of Jesus College, Oxford
and Professor of Computing Science,
Department of Computer Science

Outwitting the ‘junk news’ nation

Vidya Narayanan interviewed by Ruth Abrahams

What is junk news? How does it spread and why does it thrive on polarisation? Dr Vidya Narayanan, a researcher on the Computational Propaganda Project at the Oxford Internet Institute, talks about the key players and the AI algorithms that are gaming social media networks to promote emotionally potent, divisive political propaganda.

Ruth Abrahams: How is junk news different from news that appears in tabloid press or on populist news channels?

Vidya Narayanan: We have developed a list of five main criteria that qualify something as junk. A news source has to satisfy three of our five criteria to be junk. It's an iterative process:

- Professionalism: do these outlets refrain from providing clear information about real authors, editors, publishers and owners?
- Style: does the site use emotionally driven language with emotive expressions, hyperbole, misleading headlines, excessive capitalisation, unsafe generalisations and fallacies?
- Credibility: does the site rely on false information and conspiracy theories, and does it report without consulting multiple sources or using fact-checking methods?
- Bias: is there an ideological ‘skew’ to the site's work and does it frequently present opinion as news?
- Counterfeit: is the site mimicking real news organisations, counterfeiting fonts, branding and stylistic content strategies?

Junk news is about political, mainstream issues presented in a misleading, polarising way.

RA: Does the scale and the accuracy of junk news targeting separate it from what's gone before?

VN: Yes. Because so much public data is available online, it's possible to target people much more accurately than it was before. There are AI algorithms that are running in the background constantly building a profile of your preferences. So ‘bad actors’ then lift these techniques from the consumer industry.

RA: Have we reached a pinnacle in terms of public naivety around its own data?

VN: Awareness has increased, but it also depends on what kind of communities we're talking about. There are vulnerable communities in different parts of the world where media literacy is not so high. Coupled with a lack of technical awareness, this might still be a huge problem. The communities might not be aware

of the risks of misinformation. For audiences who are newly online I think they're encouraged to believe that this shiny new technology can't lie to us. They're predisposed to believe what they see on social media platforms.

It's a big issue. WhatsApp is a major tool for political propaganda. People tend to trust what they see in these closed networks rather than what's put out by professional news organisations.

RA: Why is social media key for these targeted political messages?

VN: AI algorithms with intent are coming together with the affordances that platforms provide – perhaps unintentionally – to promote engagement. Anything that has an emotional hook tends to get shared if you have this shock factor. There's a real intersection with the intent to manipulate with what social media is designed to do.

RA: Why does polarisation make it difficult to correct falsehoods and inaccuracy?

VN: I think it comes back to psychology. You tend to seek out news that confirms your own beliefs, which is why there can be deliberate attempts to stoke your prejudices against a certain community. You get really attached to your world view and get quite angry when it's challenged, which might be why polarisation works so well on social media, because it's emotionally charged and deliberately seeks to tell you that you are right or that you're completely wrong. There's no room for consensus. It's either black or white.

RA: What is the best way forward to mitigate against this?

VN: I think media literacy is key. Technological awareness is key. You have to engage constantly with people and not cling too closely to your own philosophy. When we put out our reports we do get to interact with a lot of alt-right media. But in some cases it's important to have a dialogue with them. I think it's very important not to judge people for their beliefs. I think that generates a lot of anger and has led to the polarisation we see today on social media.



Vidya Narayanan
Researcher, Computational Propaganda Project at the Oxford Internet Institute

Junk news is about political, mainstream issues presented in a misleading, polarising way.

Ensuring humanity has a flourishing future

If you've read articles about the existential threats facing our species, the chances are you've come across Oxford's Future of Humanity Institute (FHI).

A multidisciplinary research institute based in the University's Humanities Division, the FHI brings together the tools of mathematics, philosophy, social sciences and science to help tackle big-picture questions about humanity and its prospects.

In the words of the Institute's Director, Professor Nick Bostrom: 'At FHI we try to rise above the smoke and the din, and give some of the leading minds of the world the chance to focus on what really matters.'

The FHI has identified artificial intelligence as a key area, with particular focus on technical AI safety and the governance of AI.

The AI Safety Group at FHI is led by Dr Owain Evans and works on the long-term challenge of robust and beneficial AI systems. The potential risks from AI systems increase as they become more capable and autonomous. Dr Evans says: 'We research the foundations of systems that are robust – that is, safe in the worst case – and reliably aligned with the preferences of human overseers. The core challenge is that systems must remain robust and aligned, even if they are substantially more competent than humans. Training by reinforcement learning with a hand-coded reward function is unlikely to produce systems that remain aligned. A reward function coded by humans will have subtle loopholes that a competent agent will be able to exploit – this is known as "reward hacking".

'The goal of robustness motivates our work on the transparency and interpretability of machine learning systems. This also motivates our work on "corrigibility", or the idea that AI systems should reliably defer to humans when uncertain and allow humans to intervene in their operation at any time.'

The group's work covers a wide range of topics in AI safety, ranging from big-picture frameworks to detailed technical studies. Researchers have published on inverse reinforcement learning, safe exploration in reinforcement learning, active learning from human teachers, safe human intervention in AI systems, the safety properties of Oracle AI systems, the problem of predicting human

deliberation, and the idea of creating intelligence AI systems based on narrow specialised components rather than unitary agents. The group comprises four full-time researchers and regularly hosts interns. Its research has been published at a host of leading AI conferences and has involved collaboration with researchers and students at DeepMind, MILA (Montreal), Stanford, Berkeley, and Oxford University's Information Engineering Group.

Founded in 2017, the Governance of AI Program is directed by Professor Nick Bostrom and Professor Allan Dafoe. The programme has published numerous influential reports and papers, instigated research collaborations across the world, and advised top government officials and industry leaders on AI policy.

Professor Dafoe says: 'The Governance of AI Program is a response to this need. While AI safety research addresses the technical challenges of building safe and beneficial AI, we focus on the political contexts in which AI is built and used. Specifically, we seek to maximise the odds that the people building and using advanced AI have the goals, motivations, worldview, time, training, resources, support and organisational home necessary to do so for the benefit of humanity.'

The Governance of AI Program brings together experts in political science, international relations, security and other disciplines to undertake this research. It seeks to steer the development of AI for the common good by conducting research on important and neglected issues of AI governance, and spreading the findings of this research through policy engagement. Research falls into three main strands: politics, which focuses on domestic and international interactions between relevant actors; governance, which focuses on institutional structures; and policy, which focuses on concrete recommendations and actions. Recent research has included work on malicious use of AI, the Chinese AI landscape, surveillance, and forecasting.

To learn more about the work of the FHI, visit fhi.ox.ac.uk

The core challenge is that systems must remain robust and aligned, even if they are substantially more competent than humans.





The need for a legal framework

By Sandra Wachter

As more of our economic, social and civic interactions come to be carried out by algorithms – from credit markets and health insurance applications to recruitment and criminal justice systems – so too have concerns increased about the lack of transparency behind the technology.

AI-based systems are often opaque, hard-to-scrutinise ‘black boxes’, which leaves individuals with little understanding of how decisions are made about them.

My quest of increasing algorithmic accountability led me towards explanations: I hoped to find a legally binding right which guarantees that important algorithmic decisions that affect people’s lives have to be explained. Unfortunately my research has shown that, from a legal perspective, we have a long way to go.

Working with fellow academics Brent Mittelstadt, a data ethicist, and Chris Russell, a machine learning expert, I have been tackling the question of what good explanations might look like and whether they are technically feasible. Our recent paper on the concept of ‘counterfactual explanations’ – explaining why a negative decision has been made and how circumstances would have had to differ for a desirable outcome – has been cited, and the ideas behind it implemented, by Google.

However, our work is far from done. Although this is a major step forward, explanations of decisions are just one side of the coin in achieving true algorithmic accountability: explanation does not equal justification or legitimacy.

We know that big data and algorithms are increasingly used to assess us, predict our behaviours and preferences, and ultimately make important decisions about us. Algorithms can infer our sexual orientation, political stance and health status without our being aware. They also decide what products or newsfeeds are shown to us, as well as whether we get hired, fired or promoted, if we get a loan or insurance, among myriad other things.

Algorithms do this by drawing inferences from highly diverse and feature-rich data (such as our web browsing history or social network interactions). These inferences are often invasive, counterintuitive and non-verifiable, and we are unable to predict, understand or refute them. Yet these inferences shape our identities and reputations and steer our paths in life. Data-driven decisions thus create new opportunities for discriminatory and biased decision making.

But do data protection laws offer meaningful control over how we are being evaluated

by algorithms? Even though their purpose is to protect our private lives, EU law and jurisprudence are currently failing to protect us from the novel risks of inferential analytics. Ironically, these inferences tend to relate to low-priority personal data, which receive the least protection in law but which pose perhaps the greatest risks in terms of privacy and discrimination. Inferences are effectively ‘economy class’ personal data in Europe.

As we show in our latest paper, in standing jurisprudence the European Court of Justice (ECJ) has consistently restricted data protection law to assessment of inward personal data such as name, age or email address, rather than outward data such as inferences, opinions or assessments like credit scores. Critically, the ECJ has likewise made clear that data protection law is not intended to ensure the accuracy of decisions and decision-making processes involving personal data, nor to make these processes fully transparent.

If a person feels unfairly treated, recourse must be sought using formal procedures that are applicable to their individual case. However, very often, especially in the private sector, the way decisions are made remain within the private autonomy of the decision maker, with limited anti-discrimination regulation to govern how decisions are made and what criteria are relevant, justified and socially acceptable.

At the root of this problem is that data protection laws focus too much on the moment when data is collected but hardly at all on what happens after it has been obtained. For example, sensitive personal data on race or sexual orientation enjoys higher protection than other types, while anonymised data does not fall under data protection law at all. This stems from the idea that we can foresee potential risks and consequences when data is collected.

But this idea loses its value in the age of big data analytics. All data can be re-identified, and non-sensitive information can turn into sensitive information: postcodes, for example, can be used to infer race or sexual orientation. It is therefore time to focus on the output data and the potential impact of data analysis.

We need more focus on how, why and for

what purpose data is processed, and to work on standards for inferential analytics – such as inferring political views or (mental) health status based on browsing behaviour – that are robust and socially acceptable.

We have made several recommendations on how to close these accountability gaps and guard against the novel risks of big data and AI. These include:

- Recognition that the right to privacy is more than just ‘data protection’ – it is about identity, reputation, autonomy and informational self-determination.
- Dealing with (new) intellectual property and trade secret laws that could hinder AI transparency by providing extensive protection of commercial interests attached to the technical processes involved.
- A focus on how data is evaluated, not just collected, with a standard for the ‘right to reasonable inferences’.
- Statistically reliable data and methods for ‘high-risk’ inferences – that is, inferences that are privacy-invasive or potentially reputationally damaging, with low verifiability due to being predictive or opinion-based.

In the same way as it was necessary to create a ‘right to be forgotten’ online in a big data world, we believe it is now necessary to create a ‘right of how to be seen’. This will help us seize the full potential of AI and big data while protecting individuals and their fundamental rights.



Sandra Wachter
Lawyer and Research Fellow,
Oxford Internet Institute

Are algorithms biased?

By Helena Webb and Marina Jirotko

How fair is it that our browsing behaviours are shaped by processes we know so little about? Is it possible to design algorithms that can be fair and visible to all?

Our online lives are being shaped by processes of automated decision making.

News feeds, search engine results and product recommendations increasingly rely on algorithms to filter content and personalise what we see when we are browsing. For instance, when we enter a query into an online search engine, algorithmic processes determine the results we see and the order of those results. Similarly, when we look on Facebook and other social networks, personalisation algorithms operate to determine the adverts and posts we see in our individual accounts.

These algorithmic processes can be immensely useful. They help us cut through the mountains of information available online and direct us to those bits that are most relevant to us. However, in recent years genuine concerns have arisen that the way these algorithms operate online can lead to unfavourable, unfair or even discriminatory outcomes.

A number of public controversies have occurred, including:

- The development of an automated system to set trending news items in users' Facebook feeds in 2016. Although this was an attempt to overcome human bias in selecting news items, it was soon found that the algorithm-based system allowed false news items to be promoted alongside items containing offensive terms and images.
- The potential for personalisation mechanisms to place users within 'filter bubbles' in which they are only shown content they are already likely to like and agree with and are not challenged to consider alternative viewpoints. Following the Brexit referendum and US presidential election in 2016 there has been a great deal of debate over the extent to which these processes might limit critical thinking and vital political discussion.
- Complaints that the results of searches put into Google Images and other search engines reinforce societal prejudices – for instance by depicting black and white people differently and by portraying stereotyped gender roles. This problem might occur if the particular algorithms involved are not designed as neutral or if the datasets they are trained on are not neutral.

A further problem that exacerbates these concerns is lack of transparency. These algorithms are typically considered commercially sensitive and therefore not made available for open inspection. In any case, they are also highly technically complex and difficult for most of us to understand. How fair is it that our browsing behaviours are shaped by processes we know so little about? Is it possible to design algorithms that can be fair and visible to all?

The ongoing multi-university research project 'UnBias' recognises that the contemporary prevalence of algorithms online is an ethical issue of societal concern. We ask key questions such as: how can we be sure that algorithms are operating in our best interests? Are algorithms ever 'neutral'? And how can we judge the trustworthiness and fairness of systems that heavily rely on algorithms?

In order to answer these questions, we combine approaches from the social and computer sciences and engage with a wide range of stakeholders including industry professionals, policymakers, educators, NGOs and online users. We carry out activities to support user understanding about online environments, raise awareness among online providers about the concerns and rights of internet users, and generate debate about the 'fair' operation of algorithms in modern life.

Our project will produce policy recommendations, educational materials and a 'fairness toolkit' to promote public-civic dialogue about how algorithms shape online experiences and how issues of online unfairness might be addressed.

The EPSRC-funded UnBias project is a collaboration between the universities of Oxford, Nottingham and Edinburgh. Find out more at unbias.wp.horizon.ac.uk



Helena Webb
Senior Researcher,
Department of
Computer Science



Marina Jirotko
Professor of Human-
Centred Computing,
Department of
Computer Science

The Futuremakers podcast

Futuremakers is the fly on the wall as our academics debate key issues shaping the future of society. Season one is all about artificial intelligence...

You may have read a hundred articles about artificial intelligence and the future of society, but these longer conversations – featuring Oxford academics at the cutting edge of research and at the forefront of their professions – explore each topic in detail, from the automation of jobs to the inherent bias of algorithms.

We live in ever-changing times, so information we can trust is more important than ever before. It's not always where our academics agree that's most revealing, but where they disagree. Futuremakers invites you to those debates.

Season one available now

Listen here: po.st/futuremakers

- How will the automation of jobs likely progress?
- Are all algorithms biased?
- Is the banking sector about to change for ever?
- Is AI good for our health?
- Does AI have a gender?
- From Ada Lovelace to Alan Turing, the birth of AI
- Has AI changed the way we find the truth?
- What does AI mean for the future of humanity?
- Is China leading the way in AI?
- Season finale: AI selection box

Futuremakers design: Christian Guthrie

Podcasts Advertisers Hosting Help & Support

Sign Up Sign In

FUTURE MAKERS

UNIVERSITY OF OXFORD

Futuremakers

10 posts

Subscribe

Apple Podcasts RSS
Spotify Google Podcasts
Deezer RadioPublic

Welcome to Futuremakers, from the University of Oxford.

It's your fly-on-the-wall into our colleges, where our academics debate key issues for the future of society.

Season One is all about Artificial Intelligence...

po.st/OxfordAI

FUTURE MAKERS: Is China leading the way in AI?

FUTURE MAKERS: What does AI mean for the future of humanity?

FUTURE MAKERS: Has AI changed the way we find the truth?

FUTURE MAKERS: From Ada Lovelace to Alan Turing – the birth of AI?



UNIVERSITY OF
OXFORD

Discover #OxfordAI

Photos throughout brochure, unless marked: Pexels.com, Shutterstock, Oxford University Images.

Photo: pexels.com