Artificial intelligence: how little has to go wrong?

Autonomous weapons, driverless cars and friendly spies in the home

Summary by Stuart Parkinson

SGR's 2018 conference critically examined the rapidly expanding areas of artificial intelligence (AI) and robotics with a fascinating line-up of speakers and workshops. About 60 people attended the day. To coincide with the event, SGR launched a new briefing on the issue. The briefing and other materials from the conference can be downloaded from: http://www.sgr.org.uk/events/artificial-intelligence-how-little-has-go-wrong

AI: what are the problem areas?

Andrew Simms, Assistant Director of SGR and lead author of the briefing, was the first speaker. He gave an overview of the emerging risks of AI and robotics, and summarised the results of a survey of the views of SGR members on these issues.

A key problem, he said, was the way in which these new technologies could entrench existing problems in society and the economy. For example, he discussed the case where the company Amazon had tried to use an AI algorithm to help speed up its recruitment of new employees. But using historical data, the algorithm quickly learnt to prefer male candidates over females. In another example, a facial recognition programme being tested by a UK police force had a 98% failure rate.

Andrew also summarised the research on the ways in which AI and robotics are likely to create huge changes in employment – with several studies warning of the potential for massive job losses without concerted action by businesses and government. Developments in military and security technologies were another major concern – with particularly disturbing risks being related to nuclear weapons systems and the increasing autonomy of armed drones.

Andrew then summarised the results of the survey of SGR members. Over 80 percent believed there was a medium to high chance of things going badly wrong with AI, while 96 percent said AI needs more regulation. A particular concern was the potential for AI to help increase the power of large corporations and undemocratic governments.

Andrew concluded with a summary of key recommendations from some recent reports on AI, such as those by the international trade union, UNI Global. There was a particular focus on bans for lethal autonomous weapons and legal/ political structures that would ensure the technologies served human goals. SGR's briefing added a further recommendation: that at least 20% of R&D spending on AI be focused on understanding and preventing misuse. For more details of SGR's briefing, see p.13.

Drones, autonomy and the future of warfare

The second presentation was given by Dr Peter Burt, author of a new report on UK research and development into autonomous drones and other military systems.

Peter summarised some of the early developments in automation in military technologies – from guided missiles to robotic supply vehicles. But, he argued, it is in military drones – aerial robotic planes – that the most rapid developments are now taking place. These systems are becoming increasingly autonomous. At the moment, autonomy is restricted to functions such as take-off, self-maintenance or navigation rather than over decisions to launch weapons, but R&D – including in the UK – is increasingly moving in that direction.

One of the most frightening aspects of this field is the possibility that a crude autonomous armed drone could eventually be constructed by a small group of technologically literate people. One robotics academic recently argued that this could now be carried out "within two years".

Peter then focused on the UK situation. While the government claims that it is not developing lethal autonomous weapons, it uses a very narrow definition of this particular military technology, allowing it to fund a wide range of R&D in the area. In November 2017, it launched its national industrial strategy, making AI a priority area. The Ministry of Defence followed this with a new strategy for its R&D arm, the Defence Science and Technology Laboratory, which also emphasised AI. There are now numerous research projects involving arms corporations such as BAE Systems and Thales, as well as numerous UK universities. One particular area of interest is 'drone swarms' – where a group of small drones is used collaboratively to achieve a particular task. The focus at the moment is on their use in surveillance, but the development of swarms of armed drones is a distinct possibility.

Peter's report – published by the campaign group, Drone Wars UK – makes a number of recommendations. The first is a global ban on lethal autonomous weapons – and for the UK government to support this, which it has been reluctant to do. The report also recommends that AI research be focused on helping to understand the underlying causes of armed conflict and so help to inform strategies which could prevent it. For a longer discussion of the Drone Wars UK report, see p.8.

Robotics ethics

The next presentation was by Prof John Finney, co-author of a UN report on robotics and ethics.

John highlighted that some forms of automation have been with us for a long time – pointing to examples from the early 1900s. Some existing and planned uses were generally positive – such as landmine-clearance operations, domestic help and healthcare applications. However, it is important, he said, that we consider whether our current political and legal structures are adequate to deal with the ethical issues raised by emerging robotic technologies, especially in the military and security fields.

Existing legal instruments include international humanitarian law – which covers use of weapons and other technologies during war – and human rights law – which is generally applicable outside war.

The UN report mentioned above argues that robots can be divided into two main types – deterministic and cognitive. Deterministic robots have predictable behaviour, i.e. a human programmer can reliably predict what the robot will do if

given a set of commands. Cognitive robots, on the other hand, use machine learning and make their own decisions. Hence a programmer cannot reliably predict their behaviour. The report argued that deterministic robots can broadly be regulated within existing political and legal frameworks, but the development of cognitive robots will require new structures.

Among the recommendations of the report are: a global ban on lethal autonomous weapons; greater restrictions on the use of the current generation of military drones; ethical issues to be tackled during research and development programmes; and ethics courses as part of science and engineering degrees. For more details, see article on p.10.

Debating with a robot

For the final part of the morning session, Sarah Woods, an award-winning playwright whose works have covered AI issues, gave a rather unusual presentation. She had a conversation with Chilly the Robot via a video screen. Chilly is an experimental robot being used in Scandinavia to interact with humans in a variety of situations. Its discussion with Sarah covered the issue of robots in war. While the video conversation in this instance was scripted, it was derived from preparatory conversations between Sarah and the robot which were generated from the robot's programming. The effect on the day was intriguing, and many in the audience (myself included) were left wondering just how intelligent the robot really was.

Audience questions and discussion covered a wide range of issues. These included the difficulties of convincing senior university management of the value of ethics courses, the fallibility of AI systems, hacking of military drones, the need to prevent war rather than just controlling technologies used during war, and the response of different professions to ethical codes.

Self-driving cars

After lunch, Perry Walker of Talkshop ran an interactive workshop on the issue of self-driving cars. The audience was divided into small groups and each took one aspect of the issue to discuss, for example, road safety. Each group was given some information cards which helped them consider the issues, and the groups' thoughts were recorded on paper to be fed into a consultation by the Department of Transport. For more details, see p.12.

SGR's Annual General Meeting

The event also included SGR's AGM, chaired by Jan Maskell. SGR's Executive Director, Stuart Parkinson, presented highlights from the annual report, and the organisation's Treasurer, Alasdair Beal, presented the accounts. The National Coordinating Committee for the coming year was elected (see p.4), with the session concluding with discussion of current and planned activities.

Audience feedback on the speakers and interactive sessions was generally very positive.

The 2018 Martin Ryle Lecture, 31 October, Conway Hall, London

Research and the public good

Review by Alan Cottey

Sir Paul Nurse gave the third Martin Ryle lecture, organised by the Martin Ryle Trust and Conway Hall Ethical Society, and marking the centenary of the birth of Martin Ryle. The Trust's chair, Philip Webber, gave a brief introduction on Ryle, the Trust and the 2018 Martin Ryle lecturer. Paul Nurse is a geneticist and the director of the Francis Crick Institute, the largest single biomedical laboratory in Europe.

An organising theme in Nurse's thought about research is a division into three sectors – discovery research, translational research and applied research. He became a leading scientist in the discovery sector by studying the genetic control of cell division. It has long been obvious that this is potentially important for cancer research, and Nurse has been central in all three of the mentioned sectors. Nevertheless, it was clear from his lecture that discovery research was his first love and it has remained so, even as he moved into directing and policy roles in all of the research sectors.

Applied research at the other end of the spectrum is aimed at achieving specific outcomes. Choosing objectives well requires an understanding of the beneficiaries' needs. And the relevant knowledge base must be sufficiently well developed so that development of the application is generally foreseeable.

Between discovery research and applied research lies the 'valley of death'. Usually the focus is directly on research to bridge that gap but attention is also required on pushing the bridgeheads further out into the valley. Attempts to translate should not be premature, that is, before knowledge is sufficiently reliable.

The Crick Institute, housed in a new building in central London and informally known as Sir Paul's cathedral, has 1500 staff and 1250 scientists. Close international collaboration is important to it and to Paul Nurse's ideas. Small wonder then, that Nurse is exercised about Brexit: "Artificial barriers which reduce permeability or mutual respect between the different parts of the system, such as Brexit for example, about which I have strong views, should be resisted, as they reduce the effectiveness of the research system – both to produce knowledge and for the effective applied use of that knowledge."

Concerning the problem that research results might lead to bad consequences, Nurse's view is that this should not inhibit discovery research. Attention should be focused instead on the applications end of the research spectrum, where the objectives of the research are clearer. The question then arises – what checks can best avoid turning knowledge into harmful applications? Paul Nurse's answer is an effective, healthy democracy.

There followed a Q&A session. Here is a summary of two of them –

Q: What can we do about lack of development of new antibiotics?

A: Public intervention in the private market is needed, for example public/private partnership; also improved regulation of antibiotic use.

Q: What can be done about the government not understanding science?

A: Scientists need to be more engaged with politicians – not only when asking for funds.

In summary, over 100 participants heard a wide-ranging exposition of Sir Paul Nurse's views on the practice and organisation of scientific research, and took part in a lively discussion.