A machine intelligence commission for the UK: how to grow informed public trust and maximise the positive impact of smart machines

Geoff Mulgan

Here I make the case for creating a Machine Intelligence Commission – a new public institution to help the development of new generations of algorithms, machine learning tools and uses of big data, ensuring that the public interest is protected.

I argue that new institutions of this kind – which can interrogate, inspect and influence technological development – are a precondition for growing informed public trust. That trust will, in turn, be essential if we are to reap the full potential public and economic benefits from new technologies. The proposal draws on lessons from fields such as human fertilisation, biotech and energy, which have shown how trust can be earned, and how new industries can be grown. It also draws on lessons from the mistakes made in fields like GM crops and personal health data, where lack of trust has impeded progress.

Background

We are surrounded by algorithms making decisions about our lives – predicting fires, your chances of going to hospital in the next year, the flow of traffic, whether you deserve a loan, continuously personalising your experience of the web from the answers to your search queries to the prices of goods you are shown. It's already largely technically feasible for machine intelligence to fine you every time you speed, or adjust your insurance premia in response to real-time biometric information; or automatically to tax drivers for digital firms like Uber.

Algorithmic regulation is also already with us - for example, any new algorithm for an autopilot system in aircraft has long had to be approved by regulators. There's little doubt that algorithms are going to become more powerful, more ubiquitous and more controversial.

Our work at Nesta increasingly circles around this space. We have run events and research on the implications of algorithmic regulation, law and decision-making. We have designed and piloted tools for large-scale citizen involvement in democracy across Europe (in DCENT); applied the tools of citizen science to healthcare through projects like Dementia Citizens; used machine learning tools in research to map economic patterns, like the spread of digital firms and jobs in Tech Nation. And we are an investor in many companies at the cutting edge of using data, machine learning and analytic tools in healthcare, education and business, from Featurespace to Cogbooks.

All of these projects and programmes confirm for us the huge potential benefits that come from making more use of data, and more use of algorithms to interpret data and make predictions.

But we can also see all too clearly the lack of institutions empowered or enabled to handle the choices that come with important new projects: examples include the driverless car testbeds advocated by Nesta a few years ago and now going live in London; or the linking of genomic and other health data to be analysed with machine learning.

So, how should these new possibilities be supported and guided? How can we avoid the mistakes that have impeded or undermined other promising avenues of technological development?

The answers, I'll argue, are partly similar in structure to the ones that evolved from handling many other types of new technology – from electricity to nuclear power to food standards and television – all of which required new frames of debate and also new institutions to guarantee the public interest. In every case, new rules and new powers were needed to underpin successful markets, and successful uses. And in every case laissez faire, and self-regulation, although initially attractive to the industries concerned, soon turned out to be dead-ends that couldn't guarantee public trust and so backfired economically.

Public understanding

The starting point has to be better public understanding, and better understanding amongst decision makers, so that debates don't quickly polarise into exuberant over-enthusiasm on the one hand, and paranoia on the other. Machine learning systems are necessarily opaque. Their precise logic is often unintelligible even to experts, and the complexity of the whole process of data collection, pattern recognition and the machine's probabilistic judgements can be very hard for the public to understand.

This opacity means that trust issues are bound to be more charged. Individual cases can damage public confidence, like the recent revelations that the US National Security Agency is choosing targets for execution based on deeply flawed machine learning methods. In other fields fears are already affecting public services: attempts to use learning algorithms in child protective services in New Zealand have slowed in the face of public concerns, just as parallel concerns derailed the care.data programme here in the UK, long before it was offering any serious use of machine learning to interpret patterns.

Some of the risks of algorithmic decision-making are well known. Machine learning can encode bias into data collection and analysis, hiding discrimination behind seemingly objective numbers. Algorithms are blind to effects they have in the world beyond the things they are told to measure, and there are issues of resilience if decision algorithms are built on top of each other and rely on the same interconnected data.

We want the public to understand the potential benefits of algorithms: how algorithms can be far superior to people in predicting recidivism, making diagnoses, or assessing credit. But they also need to know how to ask the right questions, and why, for example, we shouldn't believe the data too much. Incomplete and inaccurate portraits of individuals too often become treated as truths. What's recommended for me by the supposedly smart algorithms of Amazon and Youtube, that are updated every few hours, still bears only a very thin relationship to my real desires.

Just as problematic is the view that only what can be measured can be managed (perhaps the most idiotic and most often repeated statement of the modern era, usually by people who would never be foolish enough to apply it to their own lives). It risks being amplified in an age of algorithms, when what we really need is to understand what things can be measured, predicted and subjected to computer logic, and what things cannot.

Balancing public and private, open and closed

So we are bound to need an active debate about how to handle the balance between private and public, open and closed ways of handling them. That will require both clear principles and attention to detail.

For example, it's appealing to argue that all data surrounding machine learning that affects the public interest should be open. For sure, data that is publicly generated or publicly funded should be. Certainly algorithms in fields like welfare to work, health or probation, that are paid for by taxpayers, should be as transparent as possible, and in particular training data should be open since that's what - in many cases - shapes the algorithms. But we shouldn't expect too much from transparency, which can be hard to make the most of without what are still scarce technical skills. The experience of Freedom of Information laws and open data was rather different from what many expected, primarily because of very uneven capability to use the information that was freed.

Similarly, although in principle it is appealing that all algorithms should be open source, this will not always be plausible, since it may undermine the business models and thus investment in otherwise beneficial approaches, and may also lead to problematic 'gaming' of the system. In every aspect we now need a much more nuanced, and granular debate.

Economics and digital commons

Part of that debate needs to be about the economics - how to finance the very different types of algorithmic and machine intelligence that are now possible. I wrote a paper recently about the economics of digital commons and quasi-commons. Many firms have found ways of funding what are de facto commons through advertising, or data sale. Some work on contract to governments. But there are many clearly valuable commons that have not found the right economic model, and there has been less creativity than around previous technologies in finding alternatives, particularly ones that fund commons as commons.

These as-yet unrealised models are in fields such as health, law, evidence and the media. They are technically feasible; clearly would meet unmet public needs; but cannot find a sustainable business or financial model precisely because of their nature as commons.

This is set to become a much more important issue as the potential of linking multiple data sets, and in turn linking them to smart algorithms, begins to transform fields like transport. Much of the potential gain will come from interconnection and pooling. But that will require that more of the data, and more of the algorithms used to make sense of the data (e.g. from Uber, or from mobile phone companies providing a proxy for movements), are organised as commons rather than as private commodities.

Adaptive rules

To bring some of these issues together and help us navigate the choices we now need to start designing new institutions. It's currently no-one's job to work out what needs to be done. As a result the space will be partly filled by well-intentioned private initiatives. These will be useful in their own right. But they are unlikely to have the clout or credibility to deal with the more serious potential problems. As a result there will be a growing likelihood of big errors, scandals and setbacks which will make it much harder to reap the benefits.
Getting this right matters for us all as citizens. But it also matters for our economy, which should be growing large, confident new sectors offering products and services to the world.

We will need to establish some general principles – around accountability, visibility, control – but how these are articulated will require subtlety and flexibility. Fortunately we have a political and legal environment that is probably rather better suited to this task than other countries. The US has often been let down by its messy legal system, dominated by competing judges, a sometimes deductive view of the law, and huge differences between states. We also have advantages over the often equally inflexible Roman law countries in Europe with, again, a rather linear view of legal logic.

Indeed you could say that common law at its best is like an adaptive algorithm, well-suited to rather unpredictable vectors of technological change. Our best regulatory institutions have, likewise, combined some broad principles, strong accountability, and plenty of flexibility in how they interpret their remit, with plenty of space for conversation, iteration rather than over-abstraction, and without excessive reverence for the past. These approaches, guided by explicit ethical reasoning and public dialogue, have helped us to strike sensible positions on issues such as human fertility, industrial biotech and synthetic biology.

A new institution

Applying some of these lessons to this field, work needs to begin now on the detailed options for designing a Machine Intelligence Commission to guide behaviours, understanding, norms and rules.

It would not have formal regulatory powers of approval or certification. Instead it should have strong powers of investigation, and of recommendation - much like the now disbanded Royal Commission on Environmental Pollution. Parliament would give it strong powers of access to information and software, drawing on precedents in finance. To make these powers meaningful it would have strong technical capabilities including the powers to design its own algorithms and machine learning tools to interrogate other ones, for example seeking out implicit biases.

Its work would look at key sectors – transport, employment, health, finance - in order to make recommendations to existing regulatory bodies and government departments about potential risks and abuses, for example the, Financial Conduct Authority, Highways Agency, DWP and others (it could also recommend how these techniques can be used positively by government - as an intelligent first mover).

To prioritise its work the starting point should be to focus on algorithms and machine learning tools with significant scale or reach and depth; and significant potential risk or public concern.

To work well it would need strong legal, social science and design capabilities, as well as technical capabilities in data, information architectures and business models. All of these skills will be essential if it is to analyse the whole process of the machine learning system- including data collection and linking, the training of systems, means by which the public may question decisions, and the interpretability of those answers, and the design of the interaction between human and machine on the ground, which we know is critical.

Such a MIC would primarily investigate behaviours – ie where there are reasons to be concerned about the results of algorithms. But it should also have the power to investigate processes. This is, obviously a much harder task, but unavoidable, especially where the risks are more material.

Guiding this work we'll need new protocols, that are bound to have to evolve over time - for example, how any algorithm is treated should depend on how much choice the individuals affected have; how much risk is involved, particularly over life and death issues; how much is push and how much is pull. There'll be all sorts of issues about algorithms doing nudges (again, push rather than pull) where norms will probably change over time; and we'll need a much more sophisticated debate about how to promote algorithmic diversity and pluralism and avoid the lock-in of new monopolies.

While much of the work may involve dealing with risks, part of the role of an MIC could also be to drive up quality – showing up bad design, and ill-conceived applications – and encouraging higher standards.

Evolution into the future

This description of an MIC version 1 could be implemented relatively quickly. But it should be clear from the start that future variants will be significantly different, adjusted to changing views of the balance of opportunity and risk. In time it could evolve to have its own formal powers of regulation, standard-setting, fining or directing. It may be valuable for both the government and industry to develop certification schemes which assure the quality and fairness of systems without revealing valuable IP. Lessons could be learned from sustainability certification, which has developed auditing methods for complex, dynamic systems covering both technical specifics and broad, long-term social impact, and which are now as valuable to industry safeguarding supply chains as they are to consumers seeking reassurance.

In time it will need a clear division of labour with other bodies - such as the Competition and Markets Authority to handle the economic risks – where monopolistic and predatory positions are being built up, and could easily become entrenched; and others like the Information Commissioner's Office, the Open Data Institute, Government Digital Service and others, for example reinforcing the drive for open registries of data wherever possible (a move that could be supported by other regulation such as tying university funding to open access model of data ownership).

In time it could also help to shape the arrangements for new commons – such as the management of transport flows in a city; or handling of open data in banking.

An interesting question, which will come up before long, is one of style. As critical choices become visible, the MIC is likely to need to create a strong leadership with the personality, and visibility, to begin a conversation with the public, on the evening news, about both the advantages and disadvantages of new technologies. We've learned from other fields that the most successful regulators are embodied in an individual (even when governance is formally in the hands of a committee). Their personality, their battles, and their arguments, help to build trust through narratives of exploration, argument and resolution. The Human Fertilisation and Embryology Authority for example was credible both for its technical expertise, and for its ethical and political nous, in large part because of the personality of its leading figures.

Initially the MIC should use powers of investigation to define the spaces in which more formal and visible regulation will contribute to the public interest. But it should recommend periodically to parliament what stronger powers might be needed to ensure both greater public trust and stronger grounds for public trust. What is the alternative? The alternatives are either to leave a vacuum, which is highly risky, or to rely on self-regulation and private initiative, which, even where well-intentioned, and useful in establishing greater public awareness, have a poor track record in fields like finance and the media and are very unlikely to do any better in this field where the potential benefits and risks are so much greater.

Designing a credible, strong, knowledgeable institution charged with growing informed and justifiable trust would help the UK to retain a strong position in the useful application of new tools, while also becoming a countervailing force against potential abuses of power.

In the long-run the stakes are huge. But we have time to evolve institutions to fill this space. So what I propose is a deliberately evolutionary approach, an adaptive approach to handling an extraordinary new power to the maximum public benefit, but one that should start now.

(This paper is based on a talk given to the Alan Turing Institute in London, February 2016).

Geoff Mulgan is Chief Executive of Nesta. He has worked in governments (as head of policy for Prime Minister Tony Blair and head of the UK government Strategy Unit, and as an adviser for many other governments around the world); in telecoms (in which he has a PhD, and has been an investor, funder and researcher on digital economies); as a social entrepreneur (for example, establishing the global social innovation exchange and a network of new schools); and as an author of books translated into dozens of languages. He has been a visiting professor at University College London, LSE, Melbourne University and is currently senior visiting scholar at Harvard.

info@nesta.org.uk @nesta_uk www.facebook.com/nesta.uk www.nesta.org.uk

Nesta is a registered charity in England and Wales with company number 7706036 and charity number 1144091.
Registered as a charity in Scotland number SCO42833. Registered office: 1 Plough Place, London, EC4A 1DE.

February 2016