Nesta is an innovation foundation, based in the UK. This submission contains our thoughts on algorithms in decision-making, based on our research including:

- Machines that learn in the Wild
- Wise Councils
- Machine Intelligence Commission
- Shadow of the Smart Machine series

In this submission we discuss the existing use of basic algorithms and the emergence of machine learning algorithms in decision making. We highlight a few examples and pull on innovative work at the local level using predictive analytics to support child services across the UK. We also discuss the issue of transparency and whether it really is the answer to the challenges that complex, opaque machine learning systems create for decision making. Finally we briefly touch on work Nesta is undertaking with BSI and our previous suggestions for a new Machine Intelligence Commission.

Basic algorithms vs machine learning

Firstly it is important to make a distinction between algorithms as a catch all term and the more complicated machine learning algorithms that sit behind many Artificial Intelligence (AI) applications, predictive analytics systems and more advanced decision support systems. An algorithm is a basic set of instructions (which could be as simple as a checklist or decision flow chart), in computer science these instructions can be used to run calculations and data analysis or reasoning tasks.

Advances in computer science and computer processing power have seen an explosion in the development and use of more advanced 'learning' algorithms, called machine learning algorithms, which are able to learn from data and even their environment to create and continually adapt their set of instructions to complete certain tasks or generate insights. While it is this class of algorithms that has really brought to the fore the questions the committee is asking, the existing use of basic algorithms have important implications for how these more advanced systems will be used and their impact. While much of the anxiety is around algorithms that will and perhaps already do, make decisions by themselves in reality they will largely (at least in the short to medium term) be more widely used as decision support systems by people. This interaction

\[3\] http://www.nesta.org.uk/blog/machine-intelligence-commission-uk
\[4\] http://www.nesta.org.uk/news/shadow-smart-machine
Written evidence submitted by Nesta (ALG0059)

between humans and ‘smart’ computers to make decisions presents a range of possible problems but also suggests that better designed interfaces and interaction (rather than just focussing on what the algorithm is doing inside the black box) may be enough to overcome these potential issues.

1. Past and current uses of decision algorithms

Basic algorithms (e.g. decision flow charts) and basic computer science algorithms have been used widely in and outside of government for many years to support decision making and in some cases automate decisions. For example, algorithms have been used to make school admissions decisions based on historical data since at least 1982. Decision algorithms and technology offer the potential both to standardise the decision-making process and in many cases improve its accuracy by ensuring the same array of inputs are used, in the same way, in every case.

Decision technologies were first introduced as a way of standardising and improving organisational decision-making. In banking, for example, decision analysis was developed to mitigate financial risks. These algorithms synthesised information and preferences to recommend whether a customer qualified for a loan. The airline industry was an early adopter of many of these technologies and IT to improve safety and is still one of the sectors with the highest use of decision algorithms (of all forms). In healthcare decision support tools have been used since at least the 1960s. Some of the earliest applications of decision algorithms in health care were for deciding how best to manage patients with particular diseases or conditions, such as acute renal failure or head and neck cancer.

Today the use of these algorithms to support decision making or automate certain processes is much more widespread and we are seeing the emergence of smarter systems based on algorithms that can ‘learn’. While there is a great deal of hype around the these learning algorithms they are still largely only used extensively to support decision making in a limited number of areas. This is likely to change fairly quickly as large companies and governments continue to explore the potential of these systems.

Within Government these systems are starting to used for a number of different applications, for example the Home Office is testing the use of ‘smart’ decision algorithms in the processing of passport applications and renewals. Other departments

---

5 http://europepmc.org/backend/ptpmcrender.fcgi?accid=PMC2545288&blobtype=pdf
such as the HMRC are also exploring this area but we do not have data on how many departments are using machine learning systems. Cross government work with these machine learning algorithms is being supported by the data science team within the Cabinet Office who have produced a data science ethics framework\(^8\). Many US states use decision algorithms much more extensively in government operations particularly in more ethically fraught areas like prison sentencing and justice\(^9\).

There is also a growing use of machine learning algorithms to support decision making in a local government context, from predictive policing\(^10\) and combating rodent infestations\(^11\) to child services\(^12\) and other human services\(^13\). In many ways these are some of the most interesting examples we have seen, particularly the development of predictive analytics for child services where the UK could be seen as world leader in the ethical development and deployment of these systems (more on this below).

Healthcare is another important area that will be fundamentally changed by decision algorithms and new decision support tool in everything from hospital management to diagnosis. IBM Watson already have a trial underway at the Alder Hey Hospital\(^14\) in Liverpool to transform many of the services the hospital runs using its ‘cognitive computing’ capabilities. A large number of companies have entered this space and are creating decision algorithms that will, in the not too distant future, impact patient care all the way from pre-primary\(^15\) to long term chronic care\(^16\). Automated approaches to laboratory diagnostics using machine learning have been very successful in Uganda where a shortage of lab technicians means access to quality diagnostic services is limited and technicians are over burden (and therefore more likely to make a mistake). With a 3d printed adaptor, cameraphone and machine learning algorithm Makerere University have automated the detection of TB, malaria and parasite infections in blood samples in a way that is easily scalable\(^17\).

\(^7\) https://quarterly.blog.gov.uk/2017/02/09/data-analytics-for-more-efficient-services-and-better-lives/
\(^9\) https://www.themarshallproject.org/2015/08/04/the-new-science-of-sentencing#hW2aBXSnl
\(^12\) http://www.nesta.org.uk/publications/wise-council-insights-cutting-edge-data-driven-local-government
\(^15\) https://www.ft.com/content/aefee3b8-d1d8-11e6-b06b-680c49b4b4c0
\(^16\) http://www.medicsen.net/en/
\(^17\) http://air.ug/microscopy/
Risks and opportunities

Machine learning algorithms present huge opportunities for improving decision making processes by using novel insights gleaned from large data sets (medical imaging is beginning to benefit hugely from this) or automating decision-making to create high-throughput systems. There are also a number of potential risks and problems. A number of these have already been articulated by the committee in a past inquiry (e.g. issues of bias). Instead we will focus on the issues that arise with decision support systems (both basic and smart systems) when they are designed to inform a human’s decision making process or there is some interaction between a human and machine in the decision making process.

We will discuss three potential issues; too much or too little trust, hidden preferences and accountability.

Trust; too much or too little

The overreliance of the airline industry, particularly flight personnel, on decision algorithms and computer automation is well documented. Research has shown this heavy reliance on computer automation can erode pilots’ expertise, dull their reflexes, and diminish their attentiveness. A government report on cockpit automation, compiled by an expert panel and drawing on the same FAA data, implicated automation-related problems, such as a complacency-induced loss of situational awareness and weakened hand-flying skills, in more than half of recent accidents. These observations have important implications for other sectors that are likely to adopt more decision support technology, particularly healthcare, the automotive industry and justice.

Moving from autonomous function to human control is not necessarily a smooth transition and, arguably, is not currently designed particularly well. Experiments have found that it took drivers of autonomous vehicles an average of 17 seconds to respond to takeover requests. The fairly recent Tesla crash illustrates the potential impacts of this issue for autonomous cars.

---

On the other side of the coin doctors have been consistently reluctant to use even simple decision support algorithms that are designed to help doctors overcome cognitive biases when diagnosing a patient or ensure they collect all the information needed to make a full diagnosis. This seems to be partly a cultural issue and partly an issue of professional pride but it is important to note that patients are also more uncomfortable with doctors who use some sort of computer decision support system to aid diagnosis.

Hidden preferences

While many of these decision tools seem to be objective, relying on data to make decisions, there are always preferences or judgements which have been built in which are often hidden to the end user. For example the CliniCal™ modified Alvarado tool is designed to be run as an app on a smart phone or tablet and is used by clinicians in the management of patients with suspected appendicitis. After using an algorithm to calculate the likelihood of appendicitis, based on the inputted data, the tool recommends a course of action for managing the patient, namely whether to observe or to admit. On the face of it, the tool does not appear to incorporate any preferences but in reality the algorithm incorporates certain preferences determined by the developer of the tool. For example, it recommends that surgeons should observe patients with a risk score below 7/10 but that they should admit patients with a risk score of 7/10 or above. This cut-off at the 7/10 level reflects the tool developer’s preferences for managing risk. A more cautious tool developer might have chosen a lower cut-off for this threshold, whereas a less cautious developer might have chosen a higher cut-off.

Even more troubling is the potential for gaming within these tools. Similar decision support tools exist for other conditions and determining the correct doses of certain drugs such as Warfarin. Warfarin itself is actually a very cheap drug, but there are stories of unscrupulous companies building loaded tools that have an in-built preference for prescribing more expensive treatments.

A recent review suggested that people tend to make “better” choices (i.e. better satisfaction with the decision made and better value congruence) when the preferences within a decision tool are made explicit and modifiable. Geraint Lewis, Chief Data

24 http://www.healthcarebusinesstech.com/patients-dont-like-clinical-decision-support-tools/
Written evidence submitted by Nesta (ALG0059)

Officer at NHS England, has recommended “whenever you come across a new decision technology, you should ask yourself three questions: first, are the preferences the tool uses to make its recommendations clear (remember that all decisions are made on a combination of both information and preferences)? Second, are the weightings for these preferences explicit – can you see them? Finally, can the patient and clinician adjust these preferences to suit their personal values?” 27

Accountability

Where decisions are made by both algorithms and people there is an important question around who is responsible should anything go wrong. Issues around accountability could push people to trust the machine learning algorithm less if they are ultimately responsible or potentially not question the decision algorithm as doing so may open them up to greater scrutiny. This becomes particularly relevant when there is disagreement about the right course of action. If the person is not responsible but with the algorithm, does that mean the company that created or sold the system is at fault when something goes wrong?

2. 'Good practice' in algorithmic decision-making

The case studies of Newcastle and Manchester city councils outlined in Nesta’s Wise Councils report 28 are example of good practice in algorithmic decision making. These councils and Hackney (discussed in more detail below) have developed data sharing systems and predictive analytics to better identify children in potential need of child services support or intervention. These projects followed pioneering work done in New Zealand where these approaches for child services were first tried.

The important points to note:

- Ethical reviews were undertaken to look at questions of bias within the data and the potential effects on people who would come into contact with these services 29
- Review and extensive consultation on the governance and use of data
- Testing and evaluation of the decision algorithms using historical and incoming data

- Appropriate process design; building in ethical gateways during implementation and finding the ‘sweet spot’ where the decision algorithm can input information or a decision in a way that it provides added value but does not

Hackney council example:

Hackney Council are working closely with four other councils and the company Xanthura to develop a model which helps identify families with an increased risk of ending up in statutory children’s social care within the next year and therefore should receive support from Early Help Services.

While developing the project Xanthura, along with information governance Officers and the legal teams of the four pilot Councils, developed information governance processes. This, along with extensive consultation with the Information Commissioner’s Office (ICO), has ensured that the project was implemented to the highest standards of compliance.

Xanthura also worked hard to support council and agency staff in the development and use of these analytical systems. Service users were engaged through desk based piloting of the system to help build confidence and trust in its use. ‘Ethical gateways’ were also built into every stage of the implementation process to assess appropriate use of the system. The process broadly involved establishing a series of ethical questions that needed to be considered, which were then mapped to various stages in the implementation process.

Approaches to overcoming potential bias

There are a number of technical approaches to identifying bias and disparate impact that do not necessarily require access to the algorithm if there were any IP issues. A number of these are outlined in a paper by Feldman et al (2014) as they relate to the legal framework in the US.

Question of transparency

Transparency is often discussed as the catch all solution to many of the problems created by machine learning algorithms. If we could understand exactly what the machine was doing and convey this to the user or human in the loop then we would be able to solve many of the challenges outline above and elsewhere. The General Data Protection Regulation (GDPR) already sets out the right to “meaningful information

---

30 https://arxiv.org/abs/1412.3756
Written evidence submitted by Nesta (ALG0059)

about the logic involved” in algorithmic decision making when profiling may be involved (and the decision is only made by a machine).

There are huge problems with this idea of transparency; transparency of what? To whom? How?

Transparency could be making the algorithm entirely open to anyone but with this approach we come up against not only potential IP issues but also the issue of understanding. The vast majority of the population will not be able to what the algorithm is doing as it is far to technical and in any case without the data that was used to train the machine learning algorithm this information is largely useless. This re-enforces the stipulation in the GDPR that the information should be meaningful but meaningful will differ hugely by who is needs the information for what end. This issue becomes more difficult in time sensitive situations where decisions need to be made quickly, for example when a doctor is diagnosing a patient in A&E or the driver behind the wheel of an autonomous car needs to take over control. Here it is not extra information about the logic of the algorithm that is needed, but trust in its outputs and well designed interactions between the human and machine.

Rather than seeking to make the logic completely transparent it may be enough to ensure (potentially through regulation) that any inbuilt preferences are not hidden and that the system is properly validated and evaluated, for example for healthcare the algorithm would be validated in a clinical setting. The need for transparency or understanding will vary greatly depending on context and should not be seen as a catch all solution. As Nesta Chief Executive Geoff Mulgan wrote in his proposal for a Machine Intelligence Commission; “In every aspect we now need a much more nuanced, and granular debate.”31

3. Regulatory oversight of decision algorithms

Standards have consistently been brought up as an important step in the governance and broader adoption of these systems but as of yet little work has been done in this space. Nesta is currently running a project with the British Standards Institute to explore the potential for standards as a governance tool for machine learning algorithms in different sectors. But to develop standards effectively we need to have a better contextual understanding of how the decision algorithms are being used and the practical implications of their use. This is why Nesta is also currently running a research project to more deeply explore the use of machine learning algorithms in the healthcare

31 http://www.nesta.org.uk/blog/machine-intelligence-commission-uk
Written evidence submitted by Nesta (ALG0059)

sector, with a particular focus on decision algorithms. Both these projects will publish output material over the next year.

**A Machine Intelligence Commission (extract from\(^{32}\):**

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 5 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

Written evidence submitted by Nesta (ALG0059)

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.

April 2017