The use of algorithms in decision making: RSS evidence to the House of Commons Science and Technology Select Committee inquiry

About us

The Royal Statistical Society (RSS) is a learned society and professional body for statisticians and data analysts, and a charity which promotes statistics for the public good. One of our key strategic goals is to support statistics in the public interest, and our response to this inquiry considers wider implications of using algorithmic tools for data informed decisions.

1. Summary

1.1. This inquiry brings a unique opportunity for the UK to take a leading role in thinking about the issues of governance, transparency and fairness relating to algorithms. We are already well positioned in our data and technology industry, and this is an opportunity to help set global standards.

1.2. Whilst algorithms may be accused of ‘bias’ this is often due to the pre-existing data they are trained on, which in turn is a legacy of previous human bias. Indeed, there is an opportunity to make the world fairer and less biased through using algorithms, if this is approached in the right way.

1.3. Our key recommendations to the inquiry are as follows:

- Establish an independent Data Ethics Council which can consider these issues in more depth and on an ongoing basis as technology is evolving, to provide independent advice to government, the private sector and the public, allowing industry to safely innovate as there is consensus around standards and ethics.

- Allow existing law (e.g. anti-discrimination) to develop through the courts to manage new challenges posed by algorithms. Develop the ‘right to explanation’ which is mooted in the General Data Protection Regulation, and a right to appeal significant decisions where people have been discriminated against.

- Make use of existing industry regulators to take on monitoring of outcomes from algorithms to check for bias that leads to discrimination. Conduct a review of the credit scoring industry which is relatively mature in its use of algorithms, where credit scorers appear to be able to explain their algorithms to their regulators. What lessons can be drawn for other fields?

- Develop professional standards for data science, which we note is a relatively young profession with little guidance already existing. Strong ethical training should be embedded into data science courses. Professional bodies should also take a lead on developing standards, and the RSS’s Data Science Section is willing to play its part. We are supportive of a Statement on Algorithmic Transparency and Accountability by the ACM US public policy council [1], which provides a useful starting point for thinking about standards in this area.

- Legal considerations also go beyond protecting individuals. As there are far-reaching implications for competition law, the Competition and Markets Authority ought to consider the potential anti-competitive effects arising from the independent use of pricing algorithms.

2. Full response

2.1. We support the premise for this committee's inquiry into decision making by algorithms, that 'in an increasingly digital world … the impact of algorithms is far-reaching'. Data and algorithms are fundamental to our economy and to people's day to day lives. For example, global consultancy firm McKinsey estimate that $2.8 trillion was contributed to global GDP from data flows in 2014 (compared to $2.7 trillion from flows of goods) [2]. Driving this trend are new technologies and applications, and widespread adoption of digital devices, meaning that the volume and variety of data that could be made available for analysis has exponentially grown. Automated decisions in technological systems are driven by algorithms, to far greater levels of complexity than in the past, and new methods, such as deep machine learning, are breaking new ground.

2.2. The Government Office for Science has highlighted the range of potential benefits from adopting artificial intelligence (AI) and machine learning, including making public services more efficient by anticipating demand and tailoring their provision, and making decisions more transparent [3]. There will be a need to address public concerns alongside such benefits, as even those uses that are on balance well regarded by the public, such as their use for beneficial medical and public health research, can be badly affected by loss of trust.

Mechanisms for change

2.3. Ultimately, for algorithms with important societal applications (e.g. in the labour market, for access to jobs or for appraisal of performance) we believe there should be scope for appeal by members of the public who may be badly affected, as well as scope for the organisations that use such algorithms to evaluate the decisions that were taken and on what basis. Transparent and defensible statistical outputs should ideally be the end goal of innovation in these areas. In circumstances where this is not the case (there are many commercial cases for example where it could not be), developments should have a level of explainability in mind to avoid key failures for their industry and for service users.

2.4. Beyond addressing the idea of safety (which encompasses ideas of data security, privacy, and data protection), principles for fair decisions ought to consider fairness of use, and whether outcomes from data driven decisions are reducing the biases that are inherent in society, or whether they are reflecting, amplifying and embedding them [4]. The enforceability of standards will be affected by legislation, which sets overarching precedents. Driven by the Data Protection Act and the GDPR for example, the Information Commissioner's Office has formed recommendations on the data protection implications of big data, AI and machine learning [5]. It is important for existing law (e.g. anti-discrimination) to develop through the courts to manage new challenges that are posed by algorithms.

2.5. It has been suggested that a 'right to explanation' in the new EU General Data Protection Regulation (GDPR) will make a big difference in helping people to challenge algorithmic systems if they appear to be malfunctioning or perpetuating wrongs [6]. The strength of the law in this area has been disputed by Wachter *et al* who state that 'there are several reasons to doubt both the legal existence and the feasibility of [a right to explanation]' [7]. Furthermore, we note that stating the choice of algorithm, without available training data, may make reproducing the mechanism that caused the decision impossible and not therefore contestable. We suggest that the UK's adoption of the GDPR should seek to develop the 'right to explanation' concept in a way that would more readily provide clarity on ‘fairness’ issues, and thus take the lead on this concept. This should develop the right to explanation which is mooted in the GDPR, and a right to appeal significant decisions where people have been discriminated against.

2.6. A national independent Council for Data Ethics, as was proposed to this Select Committee’s ‘Big Data Dilemma’ inquiry by the RSS, would be helpful to establish a non-prescriptive consensus regarding fairness principles. We were encouraged by the government’s recommendation that ‘a Council for Data Science Ethics should be established […] to address key ethical challenges for data science and provide technical research and thought leadership on the implications of data science across all sectors’ [8]. Such a body ought to place the UK at the forefront of debates around the use of data science, and help our industry to safely innovate, by building consensus around standards and ethics. It would be able to consider issues such as the right to explanation in more depth and on an ongoing basis as technology is evolving to provide independent advice to government, the private sector and the public. Our view is that, in order to advise freely, such a body should be independent of government, and should not have a regulatory function.

2.7. Data science is a relatively young profession, with few professional standards. These should be developed, with strong ethical training embedded into data science courses, so that they can anticipate issues including with the data that they train their algorithms on. Professional bodies should also take a lead role in developing standards, and the Data Science Section of the Royal Statistical Society is willing to help in this regard.

2.8. Many of the issues raised by this inquiry will focus upon the impacts of decision-making algorithms on individuals, yet the impacts of algorithms are much more far-reaching than this. Competition law is a further important area, where issues are raised that go beyond protection for individuals. Ezrachi and Stucke are among those that have written on this: saying that the ‘questions raised and discussed are neither futuristic nor speculative’ they point to recent prosecution of the use of price-fixing algorithms in the US [9]. As there are far-reaching implications in this area, the Competition and Markets Authority ought to consider the potential anti-competitive effects arising from the independent use of pricing algorithms.

Transparency and accountability

2.9. To inform how principles for algorithms should be applied, we need ways to see and ask about decisions. Yet on a technical level, algorithms for decision making are extremely varied in their level of complexity, explainability, and area of application. If their coding is known and fixed it should be relatively straightforward for those that develop them to explain their workings. So, for many algorithms that are used, they can be readily explained if there is sufficient openness to audit, and access to expertise. However, there are also types of algorithms for which this is not the case. The most widely acknowledged examples come from new deep learning algorithms forming ‘neural networks’. These can quickly exceed our ability to understand all their functions, yet they are already finding powerful applications for which other types of algorithms would fail. We are broadly supportive of a Statement on Algorithmic Transparency and Accountability published by the Association for Computing Machinery (ACM US Public Policy Council) [1], which provides a useful starting point for industry bodies thinking about these issues. This includes ‘data provenance’, which would enable an understanding of the meta data on which the algorithm was trained.

2.10. It should be recognised that explanation of algorithms, including new ‘deep learning’ algorithms, to sufficient breadth and depth should be formed by researchers and developers, not only by regulatory pressures. We do not think that standards of algorithmic transparency can be legislatively set, as the specifics of technology, algorithms and their application vary so much. Researchers at the UCL’s Big Data Institute have written about the variability of algorithms, and of their purpose:

2.11. “Modern machine-learning algorithms are typically designed to excel in predictive accuracy using massive volumes of data. The availability of extremely large datasets, together with modern computational power, makes this approach quite practical. However, with prediction as the endpoint, such algorithms tend to assimilate the input data and construct complex models with convoluted and interacting components. […] It thus becomes difficult to unpick specific strands of the decision-making process to understand precisely how a conclusion was reached. By contrast, traditional statistical algorithms are concerned with explanation as well as prediction, and tend to use clearly specified, often linear models, which are easier to scrutinise – although they are, on occasion, less powerful. In some cases, the impressive performance of ML algorithms can make the lack of transparency a reasonable trade-off, but this may not always be the case.” (Olhede & Rodrigues, 2017 [10])

2.12. Our research in this area points, in general, to a need for caution when widening the field of application for unexplained / partially explained data science and AI, from less regulated industries where they may have been developed, to those that require much greater explainability, and where unexplained outcomes could be severe. We see important differences in the level of pressure to explain data science and statistical approaches across different industries. Discussion of this topic in the MIT Technology Review suggests that the main purpose of requiring an explanation of deep learning algorithms should be to reduce failure in the systems that they produce [11].

2.13. Examples of stronger, regulatory approaches should lead to some broadly transferable lessons. Developments in medicine and in clinical trials, for example, have become increasingly regulated to reduce potential harm, whereas other industries such as advertising, entertainment and online social media platforms are much more lightly regulated, and might remain so. In financial services, regulation appears to have re-enforced consumers’ right to request information, particularly about their credit data and how such data have informed decisions. A set of high level principles are integral to this, including that “a firm must pay due regard to the interests of its customers and treat them fairly” and that “a firm must take reasonable care to ensure the suitability of its advice and discretionary decisions for any customer who is entitled to rely upon its judgment” [12]. Industry guidance informs public guidance, disseminated by comparison sites such as Which?. We recommend that a review of the credit scoring industry should be undertaken as it is relatively mature in its use of algorithms, and as credit scorers appear to be able to explain their algorithms to their regulators. Such a review would consider what lessons can be drawn from this for other fields.

Detecting and mitigating issues of bias: the importance of data access

2.14. One of the biggest issues surrounding new machine learning algorithms is the data that they are trained on. The strengths and weaknesses of the input data are, therefore, hugely important, and should be considered as well as the inherent logic or formulation of using an automated or analytical system to address a given problem. Olhede and Rodrigues

[10] write that “Even if we can identify the variables fed into an algorithm, data which reflect poor sampling design, unconscious bias, or which contain irrelevant correlations will have repercussions for the computed output: the algorithm can only work with the data it has.”

2.15. Having more open data and data standards, of which the Open Data Institute is one prominent advocate, is important for establishing the quality of inputs. Other mechanisms for audit, for example by sharing data for research in ‘safe haven’ settings, also need to be supported, so that private data pertaining to important outcomes can be investigated when it is in the public interest to do so [13]. There may also be important scope for more advanced research in these areas, to compare the level of bias in decisions to a counterfactual (the quality of decisions that would have been taken if the algorithmic technology were not in place). Auditing on a post-hoc basis is in either case, not trivial and may at times be technically impossible, therefore those who deploy algorithms in society should also assist by considering fairness issues from the outset – they may be supported on this by accessing training, in addition to ‘fairness’ principles. Decision-makers can also of course assist transparency for researchers, by publishing the evidence that supports their decisions.

2.16. The importance of exploring and explaining data and algorithms as they are applied is that it should improve the ability to detect failures. For example, self-driving cars, if they are to be successful, will require advanced and complex input from machine learning. The level to which algorithms should be explained should differ depending on the implications of their use, and their possible consequences. We believe that there is a great deal of potential for algorithms to be used for good across society, and that adjustment in the prevailing laws and standards should not shut down innovation which can bring lots of societal gain.

Endnotes

[1] ACM US Public Policy Council (2017) *Statement on Algorithmic Transparency and Accountability* [PDF] http://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf

[2] Box 3. Valuing cross-border data flows’ in Manyika, J. Lund, S. Bughin, J. Woetzel, J. Stamenov, K. Dhingra, D. (2016) *Digital globalization: The new era of data flows* [PDF], McKinsey Global Institute. http://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Digital%20globalization%20The%20new%20era%20of%20global%20flows/MGI-Digital-globalization-Full-report.ashx

[3] Government Office for Science (2016) *Artificial intelligence: opportunities and implications for the future of decision making* [PDF] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/566075/gs-16-19-artificial-intelligence-ai-report.pdf

[4] Devlin, H. ‘Discrimination by algorithm: scientists devise test to detect AI bias’, *Guardian*, 19 December 2016. https://www.theguardian.com/technology/2016/dec/19/discrimination-by-algorithm-scientists-devise-test-to-detect-ai-bias

[5] Information Commissioner’s Office (2017) *Big data, artificial intelligence, machine learning and data protection* [PDF] https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf

[6] Goodman, B. Flaxman, S. (2016) 'European Union regulations on algorithmic decision-making and a "right to explanation"', in Workshop on Human Interpretability in Machine Learning at the International Conference on Machine Learning. Available at: https://arxiv.org/pdf/1606.08813v3.pdf

[7] Wachter, S. Mittelstadt, B. Floridi, L. (2016) 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (December 28, 2016), Available at SSRN. https://ssrn.com/abstract=2903469

[8] Recommendation 14 in 'The big data dilemma: Government Response to the Committee's Fourth Report of Session 2015–16' [webpage], 26 April 2016. Commons Select Committee > Science and Technology. https://www.publications.parliament.uk/pa/cm201516/cmselect/cmsctech/992/99204.htm

[9] Ezrachi, A. Stucke, M.E., Artificial Intelligence & Collusion: When Computers Inhibit Competition (April 8, 2015). Oxford Legal Studies Research Paper No. 18/2015; University of Tennessee Legal Studies Research Paper No. 267. Available at SSRN: https://ssrn.com/abstract=2591874

[10] Olhede, S. Rodrigues, R. (2017) 'Fairness and transparency in the age of the algorithm', *Significance*, 5 April 2017. http://onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2017.01012.x/full

[11] Knight, W. (2017) 'The dark secret at the heart of AI', *MIT Technology Review*, 11 April 2017. https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/?set=604130

[12] Quoted from the Financial Conduct Authority Handbook 'CONC consumer credit sourcebook' [webpage] https://www.handbook.fca.org.uk/handbook/CONC/1/

[13] "Transparency can be undesirable. Processes that are adversarial, such as selecting travellers for enhanced security review, or that enjoy trade secret protection, such as consumer credit scoring, cannot be made totally public without undermining their efficacy." Kroll, J. 'Accountable Algorithms (A Provocation)', *LSE Media Policy Project Blog*, 10 February 2016. http://blogs.lse.ac.uk/mediapolicyproject/2016/02/10/accountable-algorithms-a-provocation/

Evidence submitted April 2017 by RSS's policy and research manager