House of Commons

Science and Technology Committee

Algorithms in decision-making

Fourth Report of Session 2017–19

Report, together with formal minutes relating to the report

Ordered by the House of Commons
to be printed 15 May 2018

HC 351
Published on 23 May 2018
by authority of the House of Commons

Science and Technology Committee

The Science and Technology Committee is appointed by the House of Commons to examine the expenditure, administration and policy of the Government Office for Science and associated public bodies.

Current membership

Norman Lamb MP (*Liberal Democrat, North Norfolk*) (Chair)

Vicky Ford MP (*Conservative, Chelmsford*)

Bill Grant MP (*Conservative, Ayr, Carrick and Cumnock*)

Darren Jones MP (*Labour, Bristol North West*)

Liz Kendall MP (*Labour, Leicester West*)

Stephen Metcalfe MP (*Conservative, South Basildon and East Thurrock*)

Carol Monaghan MP (*Scottish National Party, Glasgow North West*)

Damien Moore MP (*Conservative, Southport*)

Neil O'Brien MP (*Conservative, Harborough*)

Graham Stringer MP (*Labour, Blackley and Broughton*)

Martin Whitfield MP (*Labour, East Lothian*)

Powers

The Committee is one of the departmental select committees, the powers of which are set out in House of Commons Standing Orders, principally in SO No 152. These are available on the internet via www.parliament.uk.

Publication

Committee reports are published on the Committee's website at www.parliament.uk/science and in print by Order of the House.

Evidence relating to this report is published on the relevant inquiry page of the Committee's website.

Committee staff

The current staff of the Committee are: Simon Fiander (Clerk); Yohanna Sallberg (Second Clerk); Dr Harry Beeson (Committee Specialist); Martin Smith (Committee Specialist); Seb Motala (Committee Specialist); Sonia Draper (Senior Committee Assistant); Julie Storey (Committee Assistant); and Sean Kinsey (Media Officer).

Contacts

All correspondence should be addressed to the Clerk of the Science and Technology Committee, House of Commons, London SW1A 0AA. The telephone number for general inquiries is: 020 7219 2793; the Committee's e-mail address is: scitechcom@parliament.uk.

Contents

Summary

Algorithms have long been used to aid decision-making, but in the last few years the growth of 'big data' and 'machine learning' has driven an increase in algorithmic decision-making—in finance, the legal sector, the criminal justice system, education, and healthcare, as well as recruitment decisions, giving loans or targeting adverts on social media, and there are plans for autonomous vehicles to be on public roads in the UK.

The case for our inquiry was made by Dr Stephanie Mathisen from Sense about Science, who raised the question of "the extent to which algorithms can exacerbate or reduce biases" as well as "the need for decisions made by algorithms to be challenged, understood and regulated". Such issues echo our predecessor Committee's concerns during their inquiries into Big Data and Artificial Intelligence. Now, more than two years have elapsed since that Committee called for an oversight body to monitor and address such issues. Our report identifies the themes and challenges that the newly established 'Centre for Data Ethics & Innovation' should address as it begins its work.

Our report comes as the EU General Data Protection Regulation (GDPR) becomes effective and in the wake of the recent controversy centred around the algorithm used by Cambridge Analytica to target political campaign messaging—a test case which reinforces the need for effective data protection regulation.

Algorithms need data, and their effectiveness and value tends to increase as more data are used and as more datasets are brought together. The Government should play its part in the algorithms revolution by continuing to make public sector datasets available, not just for 'big data' developers but also for algorithm developers, through new 'data trusts'. The Government should also produce, maintain and publish a list of where algorithms with significant impacts are being used within Central Government, along with projects underway or planned for public service algorithms, to aid not just private sector involvement but also transparency. The Government should identify a ministerial champion to provide government-wide oversight of such algorithms, where they are used by the public sector, and to co-ordinate departments' approaches to the development and deployment of algorithms and partnerships with the private sector. The Government could do more to realise some of the great value that is tied up in its databases, including in the NHS, and negotiate for the improved public service delivery it seeks from the arrangements and for transparency, and not simply accept what the developers offer in return for data access. The Crown Commercial Service should commission a review, from the Alan Turing Institute or other expert bodies, to set out a procurement model for algorithms developed with private sector partners which fully realises the value for the public sector. The Government should explore how the proposed 'data trusts' could be fully developed as a forum for striking such algorithm partnering deals. These are urgent requirements because partnership deals are already being struck without the benefit of comprehensive national guidance for this evolving field.

Algorithms, in looking for and exploiting data patterns, can sometimes produce flawed or biased 'decisions'—just as human decision-making is often an inexact endeavour. As a result, the algorithmic decision may disproportionately affect certain groups.

The Centre for Data Ethics & Innovation should examine such algorithm biases—to identify how to improve the 'training data' they use; how unjustified correlations can be avoided when more meaningful causal relationships should be discernible; and how algorithm developer teams should be established which include a sufficiently wide cross-section of society, or of the groups that might be affected by an algorithm. The new body should also evaluate accountability tools—principles and 'codes', audits of algorithms, certification of algorithm developers, and charging ethics boards with oversight of algorithmic decisions—and advise on how they should be embedded in the private sector as well as in government bodies that share their data with private sector developers. Given the international nature of digital innovation, the Centre should also engage with other like-minded organisations in other comparable jurisdictions in order to develop and share best practice.

Transparency must be a key underpinning for algorithm accountability. There is a debate about whether that transparency should involve sharing the workings of the algorithm 'black box' with those affected by the algorithm and the individuals whose data have been used, or whether, alternatively, an 'explanation' is provided. While we acknowledge the practical difficulties with sharing data in an understandable form, the default should be that algorithms are transparent when the algorithms in question affect the public. The Centre for Data Ethics & Innovation and the ICO should examine the scope for individuals to be able to challenge the results of all significant algorithm decisions which affect them, and where appropriate to seek redress for the impacts of such decisions. Where algorithms might significantly adversely affect the public or their rights, we believe that the answer is a combination of explanation and as much transparency as possible.

Overall, the GDPR will provide helpful protections for those affected by algorithms and those whose data are subsumed in algorithm development, including more explicit consent requirements, although there remains some uncertainty about how some of its provisions will be interpreted. The challenge will be to secure a framework which facilitates and encourages innovation but which also maintains vital public trust and confidence. The Centre for Data Ethics & Innovation and the Information Commissioner's Office (ICO) should keep the operation of the GDPR under review as far as it governs algorithms, and report to Government by May 2019 on areas where the UK's data protection legislation might need further refinement. They should start more immediately with a review of the lessons of the Cambridge Analytica case. We welcome the amendments made to the Data Protection Bill which give the ICO the powers it sought in relation to its Information Notices, avoiding the delays it experienced in investigating the Cambridge Analytica case. The Government should also ensure that the ICO is adequately funded to carry out these new powers. The Government, along with the ICO and the Centre for Data Ethics & Innovation, should continue to monitor how terms and conditions rules under the GDPR are being applied to ensure that personal data is protected and that consumers are effectively informed, acknowledging that it is predominantly algorithms that use those data.

'Data protection impact assessments', required under the GDPR, will be an essential safeguard. The ICO and the Centre for Data Ethics & Innovation should encourage their publication. They should also consider whether the legislation provides sufficient powers to compel data controllers to prepare adequate impact assessments.

There are also important tasks that the Centre for Data Ethics & Innovation should address around the regulatory environment for algorithms. It should review the extent of algorithm oversight by each of the main sector-specific regulators, and use the results to guide those regulators to extend their work in this area as needed. The Information Commissioner should also make an assessment, on the back of that work, of whether it needs greater powers to perform its regulatory oversight role where sector regulators do not see this as a priority.

The Government plans to put the Centre for Data Ethics & Innovation on a statutory footing. When it does so, it should set it a requirement to report annually to Parliament on the results of its work, to allow us and others to scrutinise its effectiveness.

1 Introduction

1. Algorithms have been used to aid decision-making for centuries and pre-date computers.[1] At its core, an algorithm is a set of instructions usually applied to solve a well-defined problem. In the last few years, however, "we have witnessed an exponential growth in the use of automation to power decisions that impact our lives and societies".[2] An increase in digital data and businesses with access to large datasets, and the advent of a new family of algorithms utilising 'machine learning' and artificial intelligence (AI), has driven an increase in algorithmic decision-making (see Box 1). This has spurred huge investment into this area such as the recently announced AI sector deal "worth almost £1 billion",[3] including "£603 million in newly allocated funding".[4]

Box 1: machine learning and artificial intelligence algorithms

Although there is no single agreed definition of AI,[5] there are similarities between many of those being used.[6] Broadly, AI is "a set of statistical tools and algorithms that combine to form, in part, intelligent software" enabling "computers to simulate elements of human behaviour such as learning, reasoning and classification".[7]

Often confused with AI, 'machine learning' algorithms are a narrower subset of this technology. They describe "a family of techniques that allow computers to learn directly from examples, data, and experience, finding rules or patterns that a human programmer did not explicitly specify".[8] In contrast to conventional algorithms which are fully coded, the only instructions given to machine learning algorithms are in its objectives. How it completes these are left to its own learning.

We use the term 'machine learning algorithms' in this report, although we recognise that many use it interchangeably with 'AI algorithms'

2. The availability of 'big data' and increased computational power is allowing algorithms to identify patterns in that data.[9] The Royal Society explained that because "machine learning offers the possibility of extending automated decision-making processes, allowing a greater range and depth of decision-making without human input", the potential uses are vast and it continues to grow "at an unprecedented rate".[10] Thanks to "cheaper computing power", as Google put it, "the benefits of algorithmic decision-making will become ever more broadly distributed and […] new use cases will continue to emerge."[11]

1 Q6 [Professor Nick Jennings]
2 Upturn and Omidyar Network, *Public Scrutiny of Automated Decisions: Early Lessons and Emerging Methods*, p 3
3 HM Government, '*Tech sector backs British AI industry with multi million pound investment*', 26 April 2018
4 *Industrial Strategy Artificial Intelligence Sector Deal*, April 2018
5 Science and Technology Committee, Fifth Report of Session 2016–17, Robotics and artificial intelligence, HC 145, para 4
6 See also: Shane Legg and Marcus Hutter. "A Collection of Definitions of Intelligence", Frontiers in Artificial Intelligence and Applications, Vol.157 (2007), pp 17–24
7 Transpolitica (ROB0044) para 1.4
8 Upturn and Omidyar Network, Public Scrutiny of Automated Decisions: Early Lessons and Emerging Methods, p 9
9 Q4 [Prof Louise Amoore]
10 The Royal Society (ALG0056). See also *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation* (February 2018), p 3
11 Google (ADM0016) para 2.5

3. The range of different industries in which machine learning is already being put to use includes finance (including access to loans and insurance), the legal sector, the criminal justice system, education, and healthcare, as well as recruitment decisions and targeting adverts on social media,[12] and there are plans for driverless vehicles to be on public roads in the UK in the near future.[13] Hetan Shah from the Royal Statistical Society believed that "it is best to understand this as a ubiquitous technology and to think of it almost as a public infrastructure".[14] The Royal Academy of Engineering believes that as more data are generated, an increase in the use of machine learning algorithms will allow organisations to consider a much broader range of datasets or inputs than was previously possible, providing "an opportunity for better decision-making by combining human and machine intelligence in a smart way".[15] Algorithms driven by machine learning bring certain risks as well as benefits. The first fatal collision involving an autonomous car in March 2018 has placed these technologies under heightened scrutiny[16] and led to the suspension of self-driving car tests by Uber.[17] There has been recent controversy (currently being examined by the Digital, Culture, Media & Sport Committee's inquiry into Fake News[18]) over the use of algorithms by Cambridge Analytica to identify characteristics of Facebook users to help target political campaign messaging[19]—a test case which reinforces the need for effective data protection regulation (see Chapter 4).

4. Our predecessor Committee undertook work relevant to algorithms. It reported on 'Big Data' in 2016, examining opportunities from the proliferation of big data and its associated risks.[20] It recommended the creation of what it called a "Council of Data Ethics".[21] The Committee envisaged such a body being responsible for "addressing the growing legal and ethical challenges associated with balancing privacy, anonymisation, security and public benefit".[22] In response the Government agreed to establish such a body, which would "address key ethical challenges for data science and provide technical research and thought leadership on the implications of data science across all sectors".[23] The Committee examined in 2016 the implications of the then recently approved General Data Protection Regulation (GDPR), which will become operational in May 2018, and which is being transposed into UK law through the Data Protection Bill. Our predecessor Committee's subsequent report on 'Robotics and AI', also in 2016, reiterated its call for a data ethics council and recommended that a standing 'Commission on AI' be established at the Alan Turing Institute, to be focused on "establishing principles to govern the development and application of AI techniques, as well as advising the Government of any regulation required on limits to its progression".[24] The Alan Turing Institute wrote to the Committee later in 2016, welcoming the Committee's recommendation.[25]

12 The Royal Academy of Engineering (ALG0046); Guardian News and Media (ADM0001); Q2
13 Autumn Budget, November 2017, para 4.15
14 Q8
15 The Royal Academy of Engineering (ALG0046) para 8
16 "*Self-driving cars under scrutiny after Uber pedestrian death*", Financial Times, 20 March 2018
17 "*Uber halts self-driving car tests after death*", BBC, 20 March 2018
18 Digital, Culture, Media and Sport Committee, '*Fake News*,' accessed 04 April 2018
19 "*Facebook bans political data company Cambridge Analytica*", Financial Times, 17 March 2018
20 Science and Technology Committee, Fourth Report of Session 2015–16, *The big data dilemma* HC 468
21 *The big data dilemma*, HC 468, para 102
22 Ibid.
23 Science and Technology committee, Fifth special report of session 2015–16, *The big data dilemma: Government Response to the Committee's Fourth Report of the Session 2015–16*, HC 992, para 57
24 Science and Technology Committee, Fifth Report of Session 2016–17, *Robotics and artificial intelligence*, HC 145, para 73
25 Letter from Alan Turing Institute on a '*Commission on Artificial Intelligence*', 21 October 2016

5. The Government's response, in 2017, was that work on this front was being conducted by the Royal Society and the British Academy.[26] However, while the subsequent report from these institutions, on 'Data management and use: Governance in the 21st century', rehearsed important principles around data protection, it did not tackle algorithms more generally.[27] In 2017, the Nuffield Foundation announced its intention to establish, in partnership with other bodies, a 'Convention on Data Ethics and Artificial Intelligence' to promote and support data practices that are "trustworthy, understandable, challengeable, and accountable".[28]

6. In last year's Industrial Strategy White Paper, the Government announced the establishment of "an industry-led AI Council" supported by "a new government Office for AI", to "champion research and innovation"; take advantage of advanced data analytics; and "promote greater diversity in the AI workforce".[29] The White Paper also announced that the UK would take:

> an international leadership role by investing £9m in a new 'Centre for Data Ethics & Innovation'. This world-first advisory body will review the existing governance landscape and advise the government on how we can enable and ensure ethical, safe and innovative uses of data including AI.[30]

Margot James MP, the Minister for Digital and the Creative Industries, told the House that the proposed new Centre "will advise the Government and regulators on how they can strengthen and improve the way that data and AI are governed, as well as supporting the innovative and ethical use of that data".[31] In April 2018, the Government launched its AI Sector Deal and announced that an Interim Centre for Data Ethics & Innovation "will start work on key issues straightaway and its findings will be used to inform the final design and work programme of the permanent Centre, which will be established on a statutory footing in due course. A public consultation on the permanent Centre will be launched soon."[32]

7. **The Government's proposed Centre for Data Ethics & Innovation is a welcome initiative. It will occupy a critically important position, alongside the Information Commissioner's Office, in overseeing the future development of algorithms and the 'decisions' they make. The challenge will be to secure a framework which facilitates and encourages innovation but which also maintains vital public trust and confidence.**

8. **Many of the issues raised in this report will require close monitoring, to ensure that the oversight of machine learning-driven algorithms continues to strike an appropriate and safe balance between recognising the benefits (for healthcare and other public services, for example, and for innovation in the private sector) and the risks (for**

26 Science and Technology committee, Fifth special report of session 2016–17, *Robotics and artificial intelligence: Government Response to the Committee's Fifth Report of the Session 2016–17*, HC 896. Also see: The Royal Society, '*Data management and use: Governance in the 21st century - a British Academy and Royal Society project*', accessed 21 March 2018

27 Joint report by the British Academy and Royal Society, '*Data management and use: Governance in the 21st century*', June 2017

28 Nuffield Foundation, '*Data Ethics and Artificial Intelligence*', accessed 21 March 2018

29 Industrial Strategy, November 2017

30 Ibid.

31 Data Protection Bill Committee, 22 March 2018, col 330

32 HM Government, '*Tech sector backs British AI industry with multi million pound investment*', 26 April 2018

privacy and consent, data security and any unacceptable impacts on individuals). As we discuss in this report, the Government should ensure that these issues are at the top of the new body's remit and agenda.

9. **The Government plans to put the Centre for Data Ethics & Innovation on a statutory footing. When it does so, it should set it a requirement to report annually to Parliament on the results of its work, to allow us and others to scrutinise its effectiveness. Although the terms of the Government's proposed consultation on the Centre for Data Ethics & Innovation have yet to be announced, we anticipate our report feeding into that exercise.**

Our inquiry

10. Against the background of its earlier inquiries into Big Data and AI, our predecessor Committee also launched an inquiry into algorithms in decision-making. The case for that inquiry was made by Dr Stephanie Mathisen from Sense about Science as part of her evidence to the Committee's 'My Science Inquiry' initiative, which had sought scrutiny suggestions from the public.[33] That inquiry was launched in February 2017 but ceased when the General Election was called. We decided subsequently to continue the inquiry. We received 31 submissions (78 including those from the previous inquiry) and took oral evidence from 21 witnesses including from academics in the field, think-tanks, industry and public sector organisations using algorithms, the Information Commissioner and the Minister for the Digital and the Creative Industries, Margot James MP. In addition, we held a private, introductory seminar on algorithms in October 2017, with speakers from the Alan Turing Institute and from Facebook and SAP, a software developer.[34] We would like to thank everyone who contributed to our inquiry. In April 2018 the House of Lords Committee on AI published its report.[35] We have taken their conclusions on board where relevant to our inquiry.

11. Dr Stephanie Mathisen, in her call for an algorithms inquiry, raised the question of "the extent to which algorithms can exacerbate or reduce biases" as well as "the need for decisions made by algorithms to be challenged, understood and regulated".[36] Such issues echo our predecessor Committee's concerns, albeit then expressed in the context of Big Data and AI. It is now more than two years since that Committee called for an oversight body to monitor and address such issues. Our report is intended to identify the themes and challenges that the proposed Centre for Data Ethics & Innovation should address as it begins its work. Specifically, in Chapter 2 we look at how algorithms rely on 'data sharing' and their potential for bias and discrimination. In Chapter 3 we explore ways of achieving accountability and transparency for algorithms. In Chapter 4 we consider the regulatory environment, in the light of the Cambridge Analytica case and imminent implementation of the EU General Data Protection Regulation.

33 Science and Technology Committee, Ninth Report of Session 2016–17, *Future Programme: 'My Science Inquiry'*, HC 859, para 6

34 SAP, 'Company Information,' accessed 13 April 2018

35 House of Lords Select Committee on AI, Report of Session 2017–19, *AI in the UK: ready, willing and able?*, HC 100

36 Science and Technology Committee, Ninth Report of Session 2016–17, *Future Programme: 'My Science Inquiry'*, HC 859, para 6

2 Applications and bias

Data sharing

12. A foundation for machine learning algorithms is the data on which they are built—both their initial 'training data' (paragraph 35) and the continuing feedback data which allow some algorithms to interpret and adjust to changing scenarios. 'Big data'—drawing disparate datasets together to provide new insights—requires data to flow across organisational boundaries. Our predecessor Committee's report on Big Data expounded the "enormous benefits in prospect for the economy and for people's lives" from making public data 'open'.[37] In our current inquiry we examined the way data sharing is affecting three sectors in particular—in healthcare, criminal justice and social media.

In the health sector

13. In the context of healthcare, the Academy of Medical Sciences highlighted that "machine learning algorithms are more precise and sensitive when learning from a large, high-quality set of training data."[38] According to Dame Fiona Caldicott, the National Data Guardian, "new technologies and ways of sharing data mean that we can now gain huge benefit from the sharing of health and care data".[39] Algorithms are assisting earlier and more accurate diagnosis, supporting preventative medicine, and guiding complex treatment decisions.[40] In our recent report on Genomics, we saw the potential of genomic data, when linked with other patient-related data, to find patterns for diagnosing rare diseases and 'personalising' medicine.[41] Microsoft's 'Seeing AI' application, they told us, "enables people who are visually impaired to use a mobile app that allows them to see and hear a description of what is around them".[42] AI is being used as a 'risk assessment tool' in the field of cancer.[43] In pharmacology, it is assisting in clinical trial interpretations and simulations.[44] In epidemiology, it is being "applied to public health data to detect and track infectious disease outbreak, […] enhance medical monitoring, and to optimise demand management and resource allocation in healthcare systems".[45] The recent controversy about a "computer algorithm failure" in the NHS breast screening programme shows both the benefits and the risks of some algorithms—the system allowed an enormous number of women to be automatically invited for screening at the appropriate time, but a "coding error" also meant that women aged between 68 and 71 were missed.[46]

14. Digitalisation is a key part of the NHS's strategy to use data and algorithms to improve patient care. At present, the think-tank Reform noted, "the healthcare system is still heavily reliant on paper files and most of its IT systems are not based on open-

37 Science and Technology Committee, Fourth Report of Session 2015–16, *The big data dilemma*, HC 468, para 42
38 The Academy of Medical Sciences (ALG0055) para 5
39 National Data Guardian, *National Data Guardian 2017 report published*, 12 December 2017
40 PHG Foundation (ADM0011) para 8; Research Councils UK (ALG0074) para 10; Academy of Medical Sciences (ALG0055) para 3
41 Science and Technology Committee, Third Report of Session 2017–18, *Genomics and genome editing in the NHS*, HC 349
42 Q92 [Dr M-H. Carolyn Nguyen]
43 REACT/ REFLECT research team, University of Manchester (ADM0023) para 1
44 The Academy of Medical Sciences (ALG0055)
45 Polygeia (ALG0043)
46 HC Deb, 02 May 2018, col 315

standards".[47] In 2017, Nuance Communications, a technology firm, calculated that 43% of NHS trusts were investing in some form of artificial intelligence.[48] Polygeia, a think-tank, worried that variability in NHS digitisation will mean that some trusts lag behind others in terms of improved healthcare access.[49] Reform believed that without digitalisation, adoption of machine learning in the NHS will be "sparse".[50]

15. The pace of digitisation in the NHS is slipping behind schedule. The National Information Board envisaged in 2015 that by 2020 "all patient and care records will be digital, real time and interoperable".[51] The Wachter review concluded in 2016, however, that the "journey to integrated paperless records" by 2020 was unrealistic and should be pushed back to 2023.[52] The National Advisory Group on Health Information Technology was "very concerned that an aggressive push to digitalise the entire secondary care sector by 2020 was more likely to fail than succeed".[53] The most recent annual survey by Digital Health Intelligence found falling confidence by NHS IT leaders in being able to achieve the 2020 target for integrated digital health and care records.[54]

16. The National Data Guardian, Dame Fiona Caldicott, highlighted in her 2017 report on Genomics the urgency needed in developing "consensus on the legitimacy of data sharing in order to deliver high-quality, safe and effective genetics/genomics diagnostic services".[55] Professor Harry Hemingway of the Farr Institute of Health Informatics Research emphasised that the costs of not sharing data could be "severe".[56] The revised 'Caldicott principles',[57] published in 2013,[58] emphasised that "the duty to share information can be as important as the duty to protect patient confidentiality".[59] Reform pointed out, however, that people are generally reticent to share their data because they "do not always understand what happens to their data".[60] Following the termination of the 'care.data' patient data-sharing initiative in 2016, because of its low acceptance by patients and doctors, the National Data Guardian stipulated more stringent consent agreements and opt-outs for patients.[61] She observed recently that "the most praiseworthy attempts at innovation falter if they lose public trust".[62] We explore issues around consent further in Chapter 4.

17. The current lack of digital NHS data is slowing the development of AI algorithms. Dr Dominic King, a research scientist at DeepMind Health, a company owned by Google,[63]

47 Reform, *Thinking on its own: AI in the NHS* (January 2018), p 6
48 "*New data reveals nearly half of NHS Trusts are investing in AI for patient services*", Nuance Communications, 02 January 2017
49 Polygeia (ALG0043) para 8.1
50 Reform, *Thinking on its own: AI in the NHS* (January 2018), p 6
51 National Information Board, *Delivering the Five Year Forward View* (June 2015), p 6
52 "*Paperless 2020 "likely to fail", says Wachter review of NHS IT*", Computer Weekly, 7 September 2016
53 National Advisory Group on Health Information Technology in England, *Making IT Work: Harnessing the Power of Health Information Technology to Improve Care in England* (September 2017), p 28
54 "*Confidence in achieving NHS 2020 digitisation targets falls*", Digital Health Intelligence, 18 July 2017
55 National Data Guardian, *Developing a consensus on data sharing to support NHS clinical genetics and genomics services* (August 2017), p 3
56 Q247
57 The original Caldicott Principles were developed in 1997 following a review of how patient information was handled across the NHS. Source: *Information Governance Toolkit*, Department of Health
58 UK Caldicott Guardian Council, *A Manual for Caldicott Guardians* (January 2017)
59 Information: To share or not to share? The Information Governance Review, March 2013
60 Reform, *Thinking on its own: AI in the NHS* (January 2018), p 3
61 National Data Guardian for Health and Care, *Review of Data Security, Consent and Opt-Outs* (June 2016), p 3
62 National Data Guardian, *National Data Guardian 2017 report published*, 12 December 2017
63 DeepMind was founded in 2010. In 2015 it was acquired by Google

told us that because of uncleaned, unrepresentative and disconnected NHS datasets, it took many months to produce data in "machine readable, AI-ready format for research", before DeepMind's work was able to start on an algorithm to diagnose kidney disease. Dr King wanted to see "better education and investment in what it takes to get these datasets ready, so that they can be made available to a wide group of people".[64] The differing data codes used across NHS Trusts were seen as one hindrance to the rapid processing of data. Eleonora Harwich of Reform thought that "the standardisation of clinical codes, which are going to be replaced by a standard system" would be "a positive step forward".[65]

In the criminal justice system

18. In the criminal justice system, algorithms are being used by some police forces for facial image recognition. Big Brother Watch have raised concerns about this, including about the reliability of the technology and its potential racial bias[66] (paragraph 35). The Home Office told us in our separate inquiry on biometrics that the algorithm in these systems matched video images against a 'watch list' of wanted people, but also that police operators have to confirm the match indicated by the algorithm and "people are not arrested solely on the basis of matches made by facial recognition software".[67]

19. AI and algorithms are also being used to detect "crime hotspots"[68] and find those areas most susceptible to crime.[69] Kent Constabulary have been using a commercial 'PredPol' algorithm since 2013; "a predictive policing tool" to identify areas "where offences are likely to take place" using data on past crime patterns".[70] RUSI highlighted that a similar algorithm developed in-house by Greater Manchester Police in 2012 had been "shown to be effective at reducing burglary".[71] The UCL Jill Dando Institute of Security & Crime Science emphasised that "knowing when and where a problem is most likely is only one part of the puzzle—knowing what to do is another".[72]

20. We heard in our inquiry about how Durham Constabulary is also using algorithms to "assist decision making relating to whether a suspect could be eligible for a deferred prosecution"[73] (Box 2), as well as their wider and more controversial use in the US for decisions on bail, parole and sentencing (paragraph 38).[74] Durham Constabulary believed that AI's ability to assess risk from past behaviours is being used to get "consistency in decision making" about targeted interventions for offenders.[75] HM Inspectorate of Constabulary concluded in 2017 that the wider use of the technology used at Durham could "improve effectiveness, release officer capacity, and is likely to be cost effective".[76]

64 Qq232, 234, 235
65 Q230
66 Big Brother Watch, *Big Brother Watch Briefing for Short Debate on the use of facial recognition technology in security and policing in the House of Lords* (March 2018), p8
67 *Letter from Baroness Williams of Trafford to the Committee*, 28 March 2018
68 Oxford Internet Institute (ALG0031)
69 UCL Jill Dando Institute of Security and Crime Science (ALG0048) para 5
70 Marion Oswald and Sheena Urwin submission; See also, "*Pre-crime software recruited to track gang of thieves*", New Scientist, 11 March 2015
71 RUSI, *Big Data and Policing 2017* (September 2017), p20
72 *UCL Jill Dando Institute of Security and Crime Science*
73 Sheena Urwin, Head of Criminal Justice, Durham Constabulary (ADM0032)
74 Institute of Mathematics and its Applications (ALG0028) para 19
75 Durham Constabulary (ALG0041)
76 HM Inspectorate of Constabulary, *PEEL: Police Effectiveness 2016* (March 2017), p33

Box 2: Durham Constabulary's use of algorithms

The Harm Assessment Risk Tool (HART), designed as a result of a collaboration between Durham Constabulary and Dr Barnes of University of Cambridge, is a decision support system used to assist officers in deciding whether a suspect is eligible for deferred prosecution based on the future risk of offending.

Taking 34 different predictors—information on past criminal behaviour, age, gender and postcode—the model was 'trained' on over 100,000 custody events over a five-year period. The algorithm uses all these data to make predictions on the level of risk of reoffending

Source: Sheena Urwin, Durham Constabulary

21. The HART algorithm being piloted and evaluated by Durham Constabulary does not utilise data from other police force areas, nor indeed from national IT systems.[77] The Royal United Services Institute's 'Big Data and Policing' review in 2017 concluded that "because the system was only using Durham Police's data, offences committed in other areas would not be considered, and dangerous criminals might not be identified".[78] HM Inspectorate of Constabulary found that "most forces have not yet explored fully the use of new and emerging techniques and analysis to direct operational activity at a local level".[79] Marion Oswald, Director of the Centre for Information Rights, and Sheena Urwin of Durham Constabulary, noted that only 14% of UK police forces were using algorithmic data analysis or decision-making for intelligence work.[80] Professor Louise Amoore questioned whether there is a "place for inference or correlation in the criminal justice system"[81] since, unlike normal evidence, it cannot be cross-examined or questioned.[82] Jamie Grace from Sheffield Hallam University accepted its use but wanted "a single [independent] oversight body and regulator for the use of police databases and algorithmic analysis in criminal justice".[83]

In the web and social media sector

22. On the web and social media platforms, algorithms allow faster searches. Adverts and news can be more effectively targeted. The recent controversy over the use of algorithms by Cambridge Analytica to use Facebook users' data to help target political campaigning shows the risks associated with such applications, exacerbated in that particular case by the absence of consent for use of personal data (paragraph 83). A report from the Upturn and Omidyar Network found that people have also been adversely affected where uncompetitive practices, such as distorted filtering in search engines through "algorithmic collusion", and "automatic price fixing", have been built into algorithms.[84] In 2017 the European Commission fined Google for manipulating its algorithms to demote "rival comparison shopping services in its search results" and giving "prominent placement

77 Durham Constabulary (ALG0041)
78 RUSI, *Big Data and Policing 2017* (September 2017), p 24. See also, Big Brother Watch (ADM0012) para 12
79 HM Inspectorate of Constabulary, *PEEL: Police Effectiveness 2016* (March 2017), p33
80 Marion Oswald and Sheena Urwin (ALG0030) para 5
81 Q27
82 Q28
83 Jamie Grace, Sheffield Hallam University (ALG0003) para 1
84 Upturn and Omidyar Network, *Public Scrutiny of Automated Decisions: Early Lessons and Emerging Methods*, p28

to its own comparison shopping service".[85] The Royal Statistical Society called for the Competition and Markets Authority "to consider the potential anti-competitive effects arising from the independent use of pricing algorithms".[86]

23. The major social media platforms have cemented strong market positions by providing algorithm-based services founded on vast datasets under their control. Professor Ashiq Anjum from the University of Derby explained that "smaller organisations cannot get the same benefit because they do not have access to the same wealth of data and lack the resources to invest in the technology".[87] This was also a concern of the House of Lords' Committee on AI.[88] The consolidation of platforms flows in part from their acquisition of other social media businesses, not just to acquire more customers but also to combine datasets from different but complementary applications—combining search engine, photo-sharing and messaging services—and opening up new opportunities for more sophisticated algorithms for targeting adverts and news.[89] Following Facebook's 2014 acquisition of WhatsApp, the European Commission established that Facebook were able to match Facebook users' accounts and WhatsApp users' accounts.[90] Such synergies from merging the datasets of two companies can be a key motivation for acquisitions.[91]

Government data sharing and getting value from its data

24. Our predecessor Committee's 2016 report on Big Data welcomed the progress on making government datasets 'open' to data analytics businesses and acknowledged the "vital role" played by the Government's Digital Catapult in facilitating private sector data sharing.[92] The Committee recommended that the Government produce a framework "for pro-actively identifying data sharing opportunities to break department silos"[93] and a map to set out "how the Catapult's work and its own plans to open and share Government data could be dovetailed".[94]

25. The Government-commissioned 'AI Review' in 2017 concluded that "Government and industry should deliver a programme to develop data trusts", where data-holders and data-users can share data in a "fair, safe and equitable way".[95] The 2017 Autumn Budget subsequently announced a £75m investment "to take forward key recommendations of the independent review on AI, including exploratory work to facilitate data access through 'data trusts'."[96] The Industrial Strategy White Paper suggested that the remit for

85 European Commission, *Antitrust: Commission fines Google €2.42 billion for abusing dominance as search engine by giving illegal advantage to own comparison shopping service*, 27 June 2017
86 The Royal Statistical Society (ALG0071) para 1.3
87 Professor Ashiq Anjum, Professor of Distributed Systems, University of Derby (ADM0009), para 4
88 House of Lords Select Committee on AI, Report of Session 2017–19, *AI in the UK: ready, willing and able?*, HL 100, para 122
89 A Shared Space and a Space for Sharing project (ALG0006) para 10. See also: ICO, '*Information Commissioner updates on WhatsApp / Facebook investigation*,' accessed 7 November 2016; Nick Srnicek, Platform Capitalism, (Cambridge, 2017); "*We need to nationalise Google, Facebook and Amazon. Here's why*", The Guardian, 30 August 2017.
90 European Commission, *Mergers: Commission fines Facebook €110 million for providing misleading information about WhatsApp takeover*, 18 May 2017
91 Informatica, '*The Role of Data in Mergers and Acquisitions*,' 16 December 2016
92 Science and Technology Committee, Fourth Report of Session 2015–16, *The big data dilemma*, HC 468, para 56
93 *The big data dilemma*, HC 468, para 42
94 *The big data dilemma*, HC 468, para 56
95 Professor Dame Wendy Hall and Jérôme Pesenti, *Growing the artificial intelligence industry in the UK* (October 2017), p 46
96 Autumn Budget, November 2017, para 4.10

the planned Centre for Data Ethics & Innovation "will include engaging with industry to explore establishing data trusts to facilitate easy and secure sharing of data".[97] In our current inquiry, the Government explained that:

> The idea behind 'data trusts' is that they facilitate sharing between multiple organisations, but do so in a way that ensures that the proper privacy protections and other relevant protections are in place, that there is a governance of the data, which ensures that the voices of interested parties are represented in that governance, and that there is a fair sharing of the value that can be derived from those data. That was a recommendation in the autumn, and we are beginning work now to develop that further, with an aim of piloting data trusts in future.[98]

26. Central Government's use of AI is growing. The Government's written evidence to the Lords Committee on AI highlights the use of machine learning within HMRC "as part of a goal to automate 10 million processes by the end of 2018".[99] While the Government's submission reveals where AI is used, it is more opaque about the specific ways in which it is deployed to deliver public services. Opportunities are also being explored within the Royal Navy, the MOD, and the Cabinet Office,[100] although again it was less clear how. In the 2017 Government Transformation Strategy, "making better use of data to improve decision-making, by building and expanding data science and analytical capability across government" was set as a priority.[101] Our Government witnesses told us that holistic oversight of public sector algorithms, including the "human rights issues", did not fall under a single department.[102] Last year's Autumn Budget pledged the Government to create the "GovTech Catalyst, a small central unit based in the Government Digital Service that will give businesses and innovators a clear access point to government".[103] April's AI Sector Deal allocates £20 million to a GovTech Fund to provide "innovative solutions for more efficient public services and stimulate the UK's growing GovTech sector".[104]

27. Hetan Shah from the Royal Statistical Society told us, however, that the public sector was not taking full advantage of the "extraordinary value" of the vast amount of data it already shares with private sector algorithm developers.[105] Data and algorithms are inextricably linked[106] and "algorithms are valueless without data".[107] In our recent Genomics inquiry, Genomics England and genomics scientists explained how the value of patients' genomic data could be linked to medical and other data to provide valuable insights for diagnosing rare diseases and shaping 'personalised medicine'.[108] Because of the NHS's unique scale and patient coverage, the benefits to algorithm developers of its data more generally, particularly once digitised (paragraph 15), will be enormous.

97 Industrial Strategy, November 2017
98 Q382 [Oliver Buckley]
99 Written evidence received by the House of Lords Committee on AI, HM Government (AIC0229)
100 Written evidence received by the House of Lords Committee on AI, HM Government (AIC0229)
101 Cabinet Office and Government Digital Service, *Government Transformation Strategy 2017 to 2020* (September 2017), p 6
102 Q364 [Oliver Buckley]
103 Autumn Budget, November 2017, para 6.22
104 Industrial Strategy Artificial Intelligence Sector Deal, April 2018
105 Q18
106 Q221 [Professor Harry Hemingway]
107 Q221 [Dr Dominic King]
108 Science and Technology Committee, Third Report of Session 2017–18, Genomics and genome editing in the NHS, HC 349. See also Oral evidence taken on 01 November 2017, HC (2017–18) 349, Q24 [Sir John Bell]

28. Hetan Shah told us, however, that the public sector currently has "a lack of confidence in this area and thinks the magic lies with the private sector".[109] In 2015, the Royal Free NHS Foundation Trust signed an agreement with DeepMind Health giving the company access to 1.6 million personal identifiable records (paragraph 17), but received no monetary gain in return.[110] Hetan Shah thought that the NHS was "seduced by the magic of the algorithm company" and in future should at least seek more control over the data and their transparency:

> What [the NHS] did not realise is they were the ones with the really important thing, which is the dataset. Over time, you are going to see more private sector providers springing up who can provide algorithms, but the public sector have the magic dataset, on which they have a monopoly. When they are transacting with the private sector, they should have more confidence and should not get tied up in exclusivity contracts, and they should ask for greater transparency from the private sector providers to say, 'Open up so that you can show people what is going on with this evidence'.[111]

In Reform's recent report, 'Thinking on its own: AI in the NHS', they argue that the planned 'data trusts' could provide a means for striking agreements between industry and the NHS on how "commercial value should be generated from data". Reform recommended that the Government "should explore mutually beneficial arrangements such as profit and risk-sharing agreements". Specifically:

> The Department of Health & Social Care and the Centre for Data Ethics & Innovation should build a national framework of conditions upon which commercial value is to be generated from patient data in a way that is beneficial to the NHS. The Department of Health & Social Care should then encourage NHS Digital to work with [Sustainability & Transformation Plans[112]] and trusts to use this framework and ensure industry acts locally as a useful partner to the NHS.[113]

29. **Algorithms are being used in an ever-growing number of areas, in ever-increasing ways. They are bringing big changes in their wake; from better medical diagnoses to driverless cars, and within central government where there are opportunities to make public services more effective and achieve long-term cost savings. They are also moving into areas where the benefits to those applying them may not be matched by the benefits to those subject to their 'decisions'—in some aspects of the criminal justice system, for example, and algorithms using social media datasets. Algorithms, like 'big data' analytics, need data to be shared across previously unconnected areas, to find new patterns and new insights.**

30. ***The Government should play its part in the algorithms revolution in two ways. It should continue to make public sector datasets available, not just for 'big data' developers but also algorithm developers. We welcome the Government's proposals for a 'data***

109 Q16

110 "*Google DeepMind is giving the NHS free access to its patient monitoring app*", Business Insider, 24 June 2017

111 Q15

112 Sustainability and Transformation Plans (STPs) were announced in NHS planning guidance published in December 2015. NHS organisations and local authorities in different parts of England have developed common 'place-based plans' for the future of health and care services in their area.

113 Reform, *Thinking on its own: AI in the NHS* (January 2018), pp 39–40

trusts' approach to mirror its existing 'open data' initiatives. Secondly, the Government should produce, publish, and maintain a list of where algorithms with significant impacts are being used within Central Government, along with projects underway or planned for public service algorithms, to aid not just private sector involvement but also transparency. The Government should identify a ministerial champion to provide government-wide oversight of such algorithms, where they are used by the public sector, and to co-ordinate departments' approaches to the development and deployment of algorithms and partnerships with the private sector.

31. **Algorithms need data, and their effectiveness and value tends to increase as more data are used and as more datasets are brought together. The Government could do more to realise some of the great value that is tied up in its databases, including in the NHS, and negotiate for the improved public service delivery it seeks from the arrangements and for transparency, and not simply accept what the developers offer in return for data access.** ***The Crown Commercial Service should commission a review, from the Alan Turing Institute or other expert bodies, to set out a procurement model for algorithms developed with private sector partners which fully realises the value for the public sector. The Government should explore how the proposed 'data trusts' could be fully developed as a forum for striking such algorithm partnering deals. These are urgent requirements because partnership deals are already being struck without the benefit of comprehensive national guidance for this evolving field.***

Bias

32. While sharing data widely is likely to improve the quality of the algorithms they support, the underpinning systems also need to produce reliable and fair results—without bias. Machine learning is "application agnostic".[114] Algorithms are designed to discriminate—to tell the difference—between, for example, people, images or documents. As Professor Louise Amoore of Durham University explained, "in order for an algorithm to operate, it has to give weight to some pieces of information over others", and this bias is "intrinsic to the algorithm".[115] Durham Constabulary warned against demanding "some hypothetical perfection", and instead suggested considering "the conditions that would persist if such models were not available".[116] Dr Pavel Klimov, Chair of the Law Society's Technology and the Law Group, highlighted the importance of not turning the technology into "a weapon against ourselves", referring to the need for checks and balances.[117] Some forms of bias can nevertheless extend beyond what is acceptable. Although algorithms have the potential to "promote efficiency, consistency, and fairness", they can also "reinforce historical discrimination or obscure undesirable behaviour".[118]

33. The Alan Turing Institute told us that when automated decision-making is applied "current legislation does very little to protect individuals from being discriminated" against.[119] Where algorithms are used in the criminal justice system it is imperative that algorithms are not unfairly discriminatory. This is not always the case. We were told by the Information Commissioner that "algorithmic risk scores used in some US states" to

114 Q7 [Professor Nick Jennings]
115 Q10
116 Durham Constabulary (ALG0041)
117 Q53
118 Upturn and Omidyar Network, *Public Scrutiny of Automated Decisions: Early Lessons and Emerging Methods*, p 7
119 Alan Turing Institute (ALG0073)

determine sentencing "inaccurately classified black defendants as future criminals at almost twice the rate as white defendants, perpetuating a bias that already existed in the training data."[120] Even in relatively benign uses of algorithms such as when advertisements are displayed online, the result can be that "users of that service are being profiled in a way that perpetuates discrimination, for example on the basis of race".[121]

34. Oxford and Nottingham Universities warned that as the complexity of algorithmic applications increases, "so do the inherent risks of bias, as there is a greater number of stages in the process where errors can occur and accumulate".[122] If discrimination (of the undesirable type) is introduced, subsequent deployment can amplify the discriminatory effects.[123] Discrimination can enter the decision-making process from a variety of paths—in the use of inappropriate 'training data', a lack of data, through correlation disguised as causation, or the unrepresentativeness of algorithm development teams—and can present itself at any stage of an algorithm's lifecycle including conception, design, testing, deployment, sale, or its repurpose.[124]

Training data

35. Perhaps the biggest source of unfair bias is inappropriate 'training data'[125]—the data from which the algorithm learns and identifies patterns and the statistical rules which the algorithm applies.[126] The way that training data are selected by algorithm developers can be susceptible to subconscious cultural biases,[127] especially where population diversity is omitted from the data. The Royal Society noted that "biases arising from social structures can be embedded in datasets at the point of collection, meaning that data can reflect these biases in society".[128] A well-recognised example of this risk is where algorithms are used for recruitment. As Mark Gardiner put it, if historical recruitment data are fed into a company's algorithm, the company will "continue hiring in that manner, as it will assume that male candidates are better equipped. The bias is then built and reinforced with each decision."[129] This is equivalent, Hetan Shah from the Royal Statistical Society noted, to telling the algorithm: "Here are all my best people right now, and can you get me more of those?"[130] Microsoft told us that, as part of its 'Fairness, Accountability and Transparency in Machine Learning' initiative, computer scientists were examining how some recruitment algorithms had learned biases "based on a skewed input data".[131] During our inquiry, Professor Louise Amoore of Durham University informed us of the case of a black researcher at MIT working with facial-recognition algorithms who found that

120 Information Commissioner's Office (ALG0038)
121 ICO, *Big data, artificial intelligence, machine learning and data protection* (September 2017), p 20
122 Horizon Digital Economy Research Institute, University of Nottingham, and the Human Centred Computing group, University of Oxford (ALG0049) para 4
123 The Human Rights, Big Data and Technology Project (ALG0063) para 15
124 The Human Rights, Big Data and Technology Project (ALG0063) para 23
125 IBM (ADM0017), para 10
126 Science and Technology Committee, Fifth Report of Session 2016–17, *Robotics and artificial intelligence*, HC 145, para 5
127 IBM (ADM0017), para 10
128 The Royal Society (ALG0056) para 13
129 Mark Gardiner (ALG0068) para 4
130 Q10
131 Microsoft (ALG0072) para 20

"the most widely used algorithms did not recognise her black face".[132] Professor Amoore explained that the AI had been trained to identify the patterns in a facial geometry using predominantly white faces.[133]

36. Professor Nick Jennings from the Royal Academy of Engineering believed that algorithms are "not always well trained because people do not always understand exactly how they work or what is involved in training". Because the research in this area is still relatively undeveloped, he explained, "you end up with poorly trained algorithms giving biased results".[134] This vulnerability can be difficult to tackle when, as is increasingly the case, the process of compiling training data and the process of pattern-learning are separate endeavours. Machine learning algorithm developers can procure training data from third parties, such as data brokers, where "access to the original basis on which the data was collected is unavailable".[135] The Horizon Digital Economy Research Institute explained that "as algorithms become embedded in off-the-shelf software packages and cloud services, where the algorithm itself is reused in various contexts and trained on different data, there is no one point at which the code and data are viewed together".[136]

Insufficient data

37. As well as unrepresentative data, insufficient data can also cause discrimination. As prediction accuracy is generally linked to the amount of data available for algorithm training, incorrect assessments could be more common when algorithms are applied to groups under-represented in the training data.[137] This is a recognised issue in the personal financial credit sector where, Google told us, "a lack of good data, or poor quality, incomplete, or biased datasets […] can potentially produce inequitable results in algorithmic systems".[138] Dr Adrian Weller of the Alan Turing Institute explained that one result of this 'thin data problem' is that banks may withhold credit simply because an individual does not match the pattern of larger bank customer groups:

> It will train their algorithm based on looking for people who have at least this probability [of repaying loans]. When they do that, if they happen to be looking at a particular person who comes from a demographic where there is not much data, perhaps because there are not many people of that particular racial background in a certain area, they will not be able to get sufficient certainty. That person might be an excellent [credit] risk, but they just cannot assess it because they do not have the data.[139]

132 Q10; "*Artificial intelligence: How to avoid racist algorithms*", BBC, 14 April 2017
133 Q10
134 Q10
135 The Human Rights, Big Data and Technology Project (ALG0063) para 28
136 Horizon Digital Economy Research Institute, University of Nottingham, and the Human Centred Computing group, University of Oxford (ALG0049) para 13
137 Google Research Blog, '*Equality of Opportunity in Machine Learning*', 7 October 2016
138 Google (ADM0016) para 3.8
139 Q11

Correlation without causation

38. Bias or unfairness can arise, the Royal Society told us, when a machine learning algorithm correctly finds attributes of individuals that predict outcomes, but "in contexts where society may deem use of such an attribute inappropriate".[140] The Institute of Mathematics and its Applications gave the example of an algorithm used by the courts in Broward County, Florida, which asks: 'Was one of your parents ever sent to jail or prison?' Even if predictive, the Institute emphasised the unfairness of the inference that "a defendant deserves a harsher sentence because his father went to prison".[141]

39. The sophistication of pattern-learning means that even setting restrictions on the algorithms produced, for example to ignore protected characteristics like race, may not easily solve the problem. Machine learning systems may instead identify proxies for such characteristics. Professor Amoore explained how in the US, where algorithms had been used to predict the outcome in criminal trials, "even where race as a category was removed from the input data, the algorithm still learned characteristics, or attributes, that we might say are in breach of the Equality Act, because they use what we could call proxies for race. They learn patterns in past patterns of crime or they learn patterns in postcodes, for example."[142] Following a review of Durham Constabulary's HART algorithm, used to aid custody decisions (paragraph 21), a postcode field was removed amid concerns that it could discriminate against people from poorer areas.[143] Concerns have been expressed that other characteristics used in HART and other policing algorithms are potential sources of bias, especially where they serve as proxies for race or gender. (paragraph 41).

40. The opaque nature of the algorithm 'black box' makes its use controversial in some areas. Professor Amoore warned that there may exist "areas of our social, political or economic lives where we might want to say there is no place for algorithmic decision-making".[144] She also questioned the use of inference and correlation in the criminal justice system, and suggested that its use in the US for sentencing "constitutes a violation of due process or overt discrimination".[145] (In the UK, Durham Constabulary was using an algorithm to help determine whether a low-risk offender is suitable for 'deferred prosecution'.)[146] The risk is compounded, as Professor Amoore explained, when the algorithm's results do not allow challenge:

> Whereas with conventional tools like DNA, or a photograph, or a CCTV image, or the evidence that has been given by an eye witness, there is always the possibility of this cross-examination and the questioning: 'How did you arrive at that judgment?'" With machine learning algorithms that method is obviated.[147]

140 The Royal Society (ALG0056) para 13
141 Institute of Mathematics and its Applications (ADM0008) para 23
142 Q10
143 Wired, '*UK police are using AI to inform custodial decisions – but it could be discriminating against the poor*', 1 March 2018
144 Q27
145 Kehl, Danielle, Priscilla Guo, and Samuel Kessler. 2017. *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*, accessed 5 April 2018
146 Sheena Urwin, Head of Criminal Justice, Durham Constabulary (ADM0032)
147 Q28

41. In some of our evidence, there was a desire for algorithms within the criminal justice system to be restricted to advisory roles. Elizabeth Denham, the Information Commissioner, noted that while "there may be some red lines" there is scope for its use around sensitive areas where there is "human intervention" for decisions "around sentencing or determining parole."[148] Big Brother Watch raised a concern about the Durham HART algorithm assessing reoffending risks, in part, on the basis of a wider Experian algorithm which characterises people using metrics such as postcode, family composition and occupation, which could be discriminatory.[149] Silkie Carlo, then of Liberty,[150] told us that "where algorithms are used in areas that would engage human rights, they should be at best advisory".[151]

42. At the heart of this source of bias is a propensity to confuse 'correlation', "which is what algorithms […] can detect", with 'causality'.[152] The Information Commissioner's Office explained that "where algorithmic decisions are made based on such patterns [in the data], there is a risk that they may be biased or inaccurate if there isn't actually any causality in the discovered associations."[153] Difficulties in fully understanding a machine learning algorithm, as we discuss in Chapter 3, make it hard to even identify whether correlation without causation is being applied.

Lack of representation in the algorithm development community

43. Dr Adrian Weller from the Alan Turing Institute told us that algorithm bias can also result from employees within the algorithm software industries not being representative of the wider population.[154] Greater diversity in algorithm development teams could help to avoid minority perspectives simply being overlooked, by taking advantage of a "broader spectrum of experience, backgrounds, and opinions".[155] The US National Science and Technology Council Committee on Technology concluded in 2016 that "the importance of including individuals from diverse backgrounds, experiences, and identities […] is one of the most critical and high-priority challenges for computer science and AI".[156] Dr Weller also made the case for more representation.[157] TechUK told us:

> More must be done by Government to increase diversity in those entering the computer science profession particularly in machine learning and AI system design. This is an issue that techUK would like to see the Government's AI Review exploring and make recommendations on action that should be taken to address diversity in the UK's AI research community and industry.[158]

148 Q296
149 Big Brother watch, '*Police use Experian Marketing Data for AI Custody Decisions*', 6 April 2018
150 Also known as the National Council for Civil Liberties
151 Q49
152 Institute of Mathematics and its Applications (ALG0028) para 29
153 Information Commissioner's Office (ALG0038)
154 Q10
155 National Science and Technology Council Committee on Technology, *Preparing for the future of artificial intelligence* (October 2016), p 28
156 *Preparing for the future of artificial intelligence* (October 2016), p 27
157 Q10
158 TechUK (ADM0003) para 65

44. **Algorithms, in looking for and exploiting data patterns, can sometimes produce flawed or biased 'decisions'—just as human decision-making is often an inexact endeavour. As a result, the algorithmic decision may disproportionately discriminate against certain groups, and are as unacceptable as any existing 'human' discrimination. Algorithms, like humans, can produce bias in their results, even if unintentional. When algorithms involve machine learning, they 'learn' the patterns from 'training data' which may be incomplete or unrepresentative of those who may be subsequently affected by the resulting algorithm. That can result, for example, in race or gender discrimination in recruitment processes. The patterns that algorithms rely on may be good correlations but may not in fact show a reliable causal relationship, and that can have important consequences if people are discriminated against as a result (such as in offender rehabilitation decisions). Algorithms may have incomplete data so that, for example, some do not get favourable financial credit decisions. Algorithm developer teams may not include a sufficiently wide cross-section of society (or the groups that might be affected by an algorithm) to ensure a wide range of perspectives is subsumed in their work. These biases need to be tackled by the industries involved and, as we discuss in Chapter 4, by the regulatory environment being introduced by the GDPR, and safeguards against bias should be a critical element of the remit of the Centre for Data Ethics & Innovation.**

3 Accountability and transparency

45. To the extent that algorithms affect people and the use of personal data, there must be accountability for their application, and those affected are entitled to transparency over the results and how they are arrived at.

Accountability

46. As the Information Commissioner explained, "accountability requires someone to be responsible",[159] but where responsibility for algorithms should lie can be uncertain. Nesta highlighted the need for identifying who is responsible if anything goes wrong "where decisions are made by both algorithms and people".[160] The problem, as the European Commission has acknowledged, is that "the developer of algorithmic tools may not know their precise future use and implementation" while the individuals who are "implementing the algorithmic tools for applications may, in turn, not fully understand how the algorithmic tools operate".[161]

47. Dr Pavel Klimov of the Law Society's Technology and the Law Group was wary of placing full responsibility on the user of an algorithm because strict liability may put "innocent users […] at risk of having to answer for the losses that, on the normal application of legal principles, […] they will not be liable for".[162] Dr Adrian Weller of the Alan Turing Institute believed that we already "have an existing legal framework for dealing with situations where you need to assign accountability" and considered that "we may want to assign strict liability in certain settings, but it is going to require careful thought to make sure that the right incentives are in place to lead to the best outcome for society."[163] On the other hand, Professor Alan Winfield, Professor of Robot Ethics at the University of West England, told the House of Lords Select Committee on Artificial Intelligence that "we need to treat AI as an engineered system that is held to very high standards of provable safety", and that:

> It is the designers, manufacturers, owners and operators who should be held responsible, in exactly the same way that we attribute responsibility for failure of a motor car, for instance. If there turns out to be a serious problem, generally speaking the responsibility is the manufacturers'.[164]

48. The Royal Academy of Engineering told us that "issues of governance and accountability will need to be considered in the design and development of [algorithmic] systems so that incorrect assumptions about the behaviour of users—or designers—are avoided".[165] While the submissions to our inquiry agreed that accountability was necessary, the preferred means of achieving it varied. The Upturn and Omidyar Network

159 Q313
160 Nesta (ALG0059)
161 Council of Europe, *The human rights dimensions of automated data processing techniques (in particular algorithms) and possible regulatory implications* (October 2017), p 37
162 Q83
163 Q30
164 Oral evidence taken before the House of Lords Artificial Intelligence Committee on 17 October 2017, HL (2017–18) 100, Q20 [Professor Alan Winfield]
165 Royal Academy of Engineering (ALG0046)

reported that many ways of achieving accountability which are "fairer, more interpretable, and more auditable" are being explored but that they "remain largely theoretical today".[166] We examine the scope for some of those potential accountability mechanisms below.

Principles and codes

49. Dr Sandra Wachter of the Oxford Internet Institute emphasised that standards are a prerequisite for developing a system of accountability.[167] The Information Commissioner suggested that "codes of conduct may be drawn up by trade associations or bodies representing specific sectors in order to assist the proper application of the GDPR", before being "approved by the ICO" and "monitored by an ICO-accredited body".[168] Nesta favoured the establishment of "some general principles" which "guide behaviours, understanding, norms and rules".[169]

50. There are examples of standards and principles in the field already. The Cabinet Office has published a 'Data Science Ethics Framework' for civil servants using data in policy-making.[170] In the private sector, Amazon, DeepMind/Google, Facebook, IBM and Microsoft developed their 'Partnership on AI' in 2016 "to address opportunities and challenges with AI to benefit people and society",[171] with eight tenets which include "working to protect the privacy and security of individuals" and "striving to understand and respect the interests of all parties that may be impacted by AI advances".[172] The industry-led 'Asilomar principles' include ones addressing research funding on the ethics of AI, transparency, privacy, and shared prosperity.[173] The Association for the Advancement of Artificial Intelligence and the Association of Computing Machinery have developed professional codes of ethics for the development of computer science.[174] The Institute of Electrical and Electronics Engineers, a standard-setting body, has begun work to define "ethical concerns and technical standards related to autonomous systems".[175] An MIT Technology Review has developed five principles for algorithm developers.[176] The House of Lords Committee on AI also suggests an AI code comprising of "five overarching principles",[177] calling for "intelligibility and fairness" in AI's use "for the common good", as well as the principle that AI should not be used to "diminish the data rights or privacy of individuals, families or communities".[178]

166 Upturn and Omidyar Network, *Public Scrutiny of Automated Decisions: Early Lessons and Emerging Methods*, p 8
167 Q84
168 Information Commissioner's Office (ALG0038)
169 Nesta (ALG0059)
170 Cabinet Office, *Data Science Ethical Framework*, 19 May 2016. See also The Operational Research Society (ALG0045) para. 13
171 Microsoft (ALG0072) para 13
172 TechUK (ADM0003) para 72
173 Future of Life Institute, '*A Principled AI Discussion in Asilomar*', 17 January 2017
174 TechUK (ADM0003) para 74
175 IEEE, '*IEEE Standards Association Introduces Global Initiative for Ethical Considerations in the Design of Autonomous Systems*,' 5 April 2016; Upturn and Omidyar Network, *Public Scrutiny of Automated Decisions: Early Lessons and Emerging Methods*, p 7
176 TechUK (ADM0003) para 73
177 These principles are: 1) Artificial intelligence should be developed for the common good and benefit of humanity. 2) Artificial intelligence should operate on principles of intelligibility and fairness. 3) Artificial intelligence should not be used to diminish the data rights or privacy of individuals, families or communities. 4) All citizens have the right to be educated to enable them to flourish mentally, emotionally and economically alongside artificial intelligence. 5) The autonomous power to hurt, destroy or deceive human beings should never be vested in artificial intelligence.
178 House of Lords Select Committee on AI, Report of Session 2017–19, *AI in the UK: ready, willing and able?*, HL 100, para 417

51. Despite these efforts, however, the Upturn and Omidyar Network worried that "the use of automated decisions is far outpacing the evolution of frameworks to understand and govern them".[179] Our Government witnesses told us that they were giving consideration to the 'Asilomar principles'. While there is currently no unified framework for the private sector, they hoped that the Centre for Data Ethics & Innovation would be able to help the issues "coalesce around one set of standards".[180]

Audit and certification

52. Audit is also key to building trust in algorithms.[181] The Oxford Internet Institute explained how audit can create a "procedural record to […] help data controllers to meet accountability requirements by detecting when decisions harm individuals and groups, by explaining how they occurred, and under what conditions they may occur again".[182] The Centre for Intelligent Sensing told us that audits could "probe the system with fictitious data generated by sampling from UK demographic data, or by a company's own anonymised customer data [… and] counterfactually vary the effects".[183] Auditors could then evaluate the outputs, Google explained, "to provide an indicator of whether it might be producing negative or unfair effects".[184]

53. A challenge with machine learning algorithms, highlighted by the Institute of Mathematics and its Applications, is that "there is no guarantee that an online algorithm will remain unbiased or relevant".[185] When Google Flu Trends was launched in 2008 its use of search queries to predict the spread of flu outbreaks closely matched the surveillance data from the US Centres for Disease Control, but it was reported that it then ran into difficulties when media coverage prompted flu-related searches by people who were not ill.[186] The Institute of Mathematics believed that wholly online algorithms would need their data "updated and fully revalidated".[187] The Information Commissioner called for "data scientists to find innovative ways of building in auditability, to allow an on-going internal review of algorithmic behaviour".[188]

54. Professor Daniel Neyland from Goldsmiths University believed that certificates and third-party 'seals' for algorithms that are audited could help address "the contemporary limitations of accountability and transparency in algorithmic systems", particularly if such seals are publicised.[189] The Alan Turing Institute told us that certification or seals could be used to signify "algorithms whose design, development, and/or deployment have produced fair, reasonable, and non-discriminatory outcomes".[190] The GDPR provides for certification of algorithms in terms of their privacy protections, and the Information

179 Upturn and Omidyar Network, *Public Scrutiny of Automated Decisions: Early Lessons and Emerging Methods*, p 7
180 Q375 [Oliver Buckley]
181 TechUK (ADM0003) para 4
182 Oxford Internet Institute (ALG0031)
183 Centre for Intelligent Sensing (ALG0036)
184 Google (ADM0016) para 3.17
185 Institute of Mathematics and its Applications (ADM0008) para 38
186 "*When Google got flu wrong*", Nature news, 13 February 2013
187 Institute of Mathematics and its Applications (ADM0008) para 38
188 Information Commissioner's Office (ALG0038)
189 Professor Daniel Neyland, Goldsmiths (ALG0027)
190 Alan Turing Institute (ALG0073)

Commissioner believed that "seals can help to inform people about the data protection compliance of a particular product or service". The ICO was "currently looking into how certification schemes can be set up and managed in practice".[191]

Ethics boards

55. Ethics boards can be used "to apply ethical principles and assess difficult issues that can arise in the creation and use of algorithms in decision-making", the Information Commissioner's Office told us, and can aid transparency by publishing their deliberations so that "the development of the algorithm is openly documented".[192] TechUK cautioned, however, that ethics boards "could be seen as a burden for UK SMEs that stand to benefit the most from automated decision-making technologies".[193]

56. **Setting principles and 'codes', establishing audits of algorithms, introducing certification of algorithms, and charging ethics boards with oversight of algorithmic decisions, should all play their part in identifying and tackling bias in algorithms. With the growing proliferation of algorithms, such initiatives are urgently needed. *The Government should immediately task the Centre for Data Ethics & Innovation to evaluate these various tools and advise on which to prioritise and on how they should be embedded in the private sector as well as in government bodies that share their data with private sector developers. Given the international nature of digital innovation, the Centre should also engage with other like-minded organisations in other comparable jurisdictions in order to develop and share best practice.***

Transparency

57. Algorithm accountability is often framed in terms of openness and transparency, and the ability to challenge and scrutinise the decisions reached using algorithms.[194] Although all of the details are not yet available of the recent NHS breast screening programme failure, where women aged between 68 and 71 were not sent screening appointments, it is possible that if the flaw was a relatively straightforward "coding error", as the Health Secretary put it,[195] then making that algorithm coding more widely available might have allowed the error to have been spotted much sooner. Transparency would be more of a challenge, however, where the algorithm is driven by machine learning rather than fixed computer coding. Dr Pavel Klimov of the Law Society's Technology and the Law Group explained that, in a machine learning environment, the problem with such algorithms is that "humans may no longer be in control of what decision is taken, and may not even know or understand why a wrong decision has been taken, because we are losing sight of the transparency of the process from the beginning to the end".[196] Rebecca MacKinnon from think-tank New America has warned that "algorithms driven by machine learning quickly become opaque even to their creators, who no longer understand the logic being followed".[197] Transparency is important, but particularly so when critical consequences

191 Information Commissioner's Office (ALG0038)
192 Information Commissioner's Office (ALG0038)
193 TechUK (ADM0003) para 76
194 Neyland, D. 2015. "*Bearing accountable witness to the ethical algorithmic system*", Science, Technology and Human Values, Vol.41 (2016), pp 50–76
195 HC Deb, 2 May 2018, col 315
196 Q46
197 Mark Gardiner (ALG0068) but this is referencing a quote made by Rebecca MacKinnon, director of the Ranking Digital Rights project at New America

are at stake. As the Upturn and Omidyar Network have put it, where "governments use algorithms to screen immigrants and allocate social services, it is vital that we know how to interrogate and hold these systems accountable".[198] Liberty stressed the importance of transparency for those algorithmic decisions which "engage the rights and liberties of individuals".[199]

58. Transparency, Nesta told us, could lead to greater acceptability of algorithmic decisions.[200] But transparency can take different forms—how an algorithmic decision is arrived at, or visibility of the workings inside the 'black box'. The Human Rights, Big Data and Technology Project suggested that transparency needs "to be considered at each stage in the algorithmic decision-making process, and in the process as a whole".[201] Several submissions indicated that the users of algorithms should be able to explain their decisions in terms that users can understand.[202]

59. Transparency inside the 'black box' may be of practical use only to some because, as Dr M-H. Carolyn Nguyen of Microsoft put it, it "takes a lot of data scientists to understand exactly what is going on".[203] And even then, Dr Janet Bastiman told us, "given the complex nature of these decision-making algorithms, even if the full structure, weighting, and training data were published for an end-user, it is unlikely that they would be able to understand and challenge the output from the algorithm".[204] Where algorithms are based on machine learning, Professor Louise Amoore of Durham University wondered whether full transparency was possible "even to those who have designed and written them".[205] Even if such difficulties could be overcome, University College London warned that "a central tension with making algorithms completely open is that many are trained on personal data, and some of this private data might be discoverable if we release the algorithmic models".[206]

60. Hetan Shah of the Royal Statistical Society nevertheless highlighted the recent attempts by New York City Council to require the code for all city agencies' algorithms to be published.[207] Professor Nick Jennings of the Royal Academy of Engineering, however, drew attention to the issue of 'adversarial machine learning' where individuals "know the way a machine-learning algorithm works and so you try to dupe it to believe something and come to a particular set of conclusions; then you can exploit the fact that you know that it has been mis-trained".[208] When Google originally published its PageRank algorithm nearly 20 years ago, for example, spammers gamed the search algorithm by paying each other for links and so undermined the algorithm's effectiveness.[209]

61. The Alan Turing Institute told us that "two of the biggest hurdles to a 'right of explanation' (paragraph 62) are trade secrets and copyright concerns".[210] While patents

198 Upturn and Omidyar Network, *Public Scrutiny of Automated Decisions: Early Lessons and Emerging Methods*, p 3
199 Liberty (ALG0070)
200 Nesta (ALG0059)
201 Human Rights, Big Data and Technology Project (ALG0063) para 25
202 Projects by IF (ALG0033)
203 Q112
204 Dr Janet Bastiman (ALG0029)
205 Professor Louise Amoore, Durham University, (ALG0042) para 2.3
206 University College London (ALG0050)
207 Q15 (subsequently amended to the creation of a task force to provide recommendations on how automated decision systems may be shared by the public)
208 Q19
209 Peter A. Hamilton, *Google-bombing: Manipulating the PageRank Algorithm*, p 3
210 Alan Turing Institute (ALG0073)

have traditionally been used to balance the interests of society and inventors, academics at Oxford and Nottingham universities questioned how that balance might be struck in the age of machine learning.[211] Microsoft told us that it wanted the Government to "broaden the UK's copyright exception on text and data mining, bringing it into line with that of the USA, Japan and Canada, and ensuring that the UK is well placed to be at the forefront of data analytics".[212] "Dr Janet Bastiman worried that "Since the intellectual property in machine learned systems is encapsulated in the structure, weighting, and input data that comprise the final algorithm, any legislation requiring clear transparency of the algorithm itself could have negative impact on the commercial viability of private sector institutions using this technology."[213] As Future Advocacy has recently explained, this may result in less accurate algorithms as designers opt for less accurate but easier to explain models—a concern where this affects healthcare algorithms.[214] Others cautioned against "letting commercial interests supersede the rights of people to obtain information about themselves".[215] The Upturn and Omidyar Network pointed out that France is the only country that has explicitly required disclosure of the source code of government-developed algorithms, under its open record laws.[216]

The right to explanation

62. While many of our submissions advocated a 'right to explanation',[217] the Royal Statistical Society did not think that wider "standards of algorithmic transparency can be legislatively set, as the specifics of technology, algorithms and their application vary so much".[218] The think-tank Projects by IF emphasised that "transparency is more useful with context"[219] and, comparing industries, the Royal Statistical Society found "important differences in the level of pressure to explain data science and statistical approaches".[220] Projects by IF concluded that "how a service explains its workings to users will be different to how it explains its workings to auditors."[221]

63. We heard about various ways, some in use and others in development, of facilitating a 'right to explanation'. Hetan Shah saw scope in 'counterfactual explanations';[222] an approach that Dr Wachter told us avoids having to open the black box.[223] She gave as an example where a loan application is rejected and the algorithm "would tell you what would have needed to be different in order to get the loan and give you some grounds to contest it

211 Horizon Digital Economy Research Institute, University of Nottingham, and the Human Centred Computing group, University of Oxford (ALG0049) para 19
212 Microsoft (ALG0072) para 14; The current copyright exception permits researchers with legal access to a copyrighted work, to make copies "for the purpose of computational analysis" allowing the use of "automated analytical techniques to analyse text and data for patterns, trends and other useful information". However this only exists for non-commercial research, restricting companies like Microsoft from commercialising their algorithm (Intellectual Property Office, *Guidance, Exceptions to Copyright* (November 2014))
213 Dr Janet Bastiman (ALG0029)
214 Future Advocacy & the Wellcome Trust, Ethical, Social, and Political Challenges of Artificial Intelligence in Health (April 2018), p 32
215 Horizon Digital Economy Research Institute, University of Nottingham, and the Human Centred Computing group, University of Oxford (ALG0049) para 18
216 Upturn and Omidyar Network, *Public Scrutiny of Automated Decisions: Early Lessons and Emerging Methods*, p 24
217 Aire (ALG0066); University College London (ALG0050) para 28; Alan Turing Institute (ALG0073); The Royal Statistical Society (ALG0071) para 2.5; et al
218 The Royal Statistical Society (ALG0071)
219 Projects by IF (ALG0033)
220 The Royal Statistical Society (ALG0071)
221 Projects by IF (ALG0033)
222 Q15
223 Q60

[…] This might be that if your income were £15,000 higher you would have been accepted."[224] Google thought that better 'data visualisation tools' could also help, showing "key metrics relating to an algorithm's functioning without going into the full complexity, akin to the way that car dashboards have gauges for speed, oil pressure, and so on".[225] TechUK similarly highlighted 'interactive visualisation systems'.[226] In 2017 the US Department of Defence announced funding for thirteen projects examining different approaches to making algorithms more transparent, including through visualisation tools.[227] Machine learning might in the future also be used itself to explain other algorithms.[228]

64. Whatever form transparency takes, Projects by IF emphasised that "services based on the outcome of an algorithm need to empower users to raise a complaint" if a decision is in dispute.[229] Oxford Internet Institute believed that "the rapid spread of automated decision-making into sensitive areas of life, such as health insurance, credit scoring or recruiting, demands that we do better in allowing people to understand how their lives are being shaped by algorithms".[230] IBM thought it was important that explanations uncover how algorithms "interpreted their input" as well as "why they recommended a particular output".[231]

65. Oxford Internet Institute highlighted that the 'right to explanation' is omitted from the GDPR's Article 22 (paragraph 73), and is only included in a non-legally binding recital, which serves only as guidance.[232] University College London wanted a meaningful 'right to explanation' strengthened, to include 'semi-automated' as well as the 'automated' decisions that are covered by the GDPR.[233] A 'right to information' in the Data Protection Bill gives the data subject the right "to obtain from the [data] controller, on request, knowledge of the reasoning underlying the processing" of any decision, but only in connection with intelligence services data processing. The Bill has no wider 'right to explanation' for the UK, nor one that could be applied to all decisions rather than just to the intelligence field. In France, digital-economy minister Mounir Mahjoubi recently said that its government should not use any algorithm whose decisions cannot be explained.[234]

66. **Transparency must be a key underpinning for algorithm accountability. There is a debate about whether that transparency should involve sharing the workings of the algorithm 'black box' with those affected by the algorithm and the individuals whose data have been used, or whether (because such information will not be widely understood) an 'explanation' is provided. *Where disclosure of the inner workings of privately-developed public-service algorithms would present their developers with commercial or personal-data confidentiality issues, the Government and the Centre***

224 Q60
225 Google (ADM0016) para 3.17
226 TechUK (ADM0003) para 81
227 TechUK (ADM0003); "*The U.S. Military Wants Its Autonomous Machines to Explain Themselves*", MIT Technology Review, 14 March 2017
228 A recent experiment aimed at explaining an AI system involved running another AI system in parallel, which monitored patterns in people narrating their experiences of playing a computer game. These patterns in the human explanations were learnt by the parallel AI system, and then applied to provide their own explanations. See Osbert Bastani, Carolyn Kim and Hamsa Bastani, "*Interpretability via Model Extraction*"; and "The unexamined mind", The Economist, 17 February 2018.
229 Projects by IF (ALG0033) para 6.3
230 Oxford Internet Institute (ALG0031)
231 IBM (ADM0017) para 5
232 Oxford Internet Institute (ALG0031)
233 University College London (ALG0050) para 28
234 "Humans may not always grasp why AIs act. Don't panic", The Economist, 15 February 2018

for Data Ethics & Innovation should explore with the industries involved the scope for using the proposed 'data trust' model to make that data available in suitably de-sensitised format. While we acknowledge the practical difficulties with sharing an 'explanation' in an understandable form, the Government's default position should be that explanations of the way algorithms work should be published when the algorithms in question affect the rights and liberties of individuals. That will make it easier for the decisions produced by algorithms also to be explained. The Centre for Data Ethics & Innovation should examine how explanations for how algorithms work can be required to be of sufficient quality to allow a reasonable person to be able to challenge the 'decision' of the algorithm—an issue we explore further in Chapter 4. Where algorithms might significantly adversely affect the public or their rights, we believe that the answer is a combination of explanation and as much transparency as possible.

67. ***The 'right to explanation' is a key part of achieving accountability. We note that the Government has not gone beyond the GDPR's non-binding provisions and that individuals are not currently able to formally challenge the results of all algorithm decisions or where appropriate to seek redress for the impacts of such decisions. The scope for such safeguards should be considered by the Centre for Data Ethics & Innovation and the ICO in the review of the operation of the GDPR that we advocate in Chapter 4.***

4 The Centre for Data Ethics & Innovation, research and the regulatory environment

68. The Government, after some false starts, has now made a commitment to establish an oversight and ethics body—the planned 'Centre for Data Ethics & Innovation' (paragraph 6). Many submissions to our inquiry identified a need for continuing research, which might be a focus for the work of the new body. The Royal Society, like our predecessor Committee, argued that "progress in some areas of machine learning research will impact directly on the social acceptability of machine learning applications". It recommended that research funding bodies encourage studies into "algorithm interpretability, robustness, privacy, fairness, inference of causality, human-machine interactions, and security".[235] University College London advised that the Government should invest in "interdisciplinary research around how to achieve meaningful algorithmic transparency and accountability from social and technical perspectives".[236] The think tank Future Advocacy wanted more Government research on transparency and accountability and supported more 'open data' initiatives (paragraph 24).[237] TechUK suggested that, what it called, a 'UK Algorithmic Transparency Challenge' should be created to "encourage UK businesses and academia to come up with innovative ways to increase the transparency of algorithms".[238] In April, the Government announced plans to spend £11 million on research projects "to better understand the ethical and security implications of data sharing and privacy breaches".[239]

69. **We welcome the announcement made in the AI Sector Deal to invest in research tackling the ethical implications around AI.** ***The Government should liaise with the Centre for Data Ethics & Innovation and with UK Research & Innovation, to encourage sufficient UKRI-funded research to be undertaken on how algorithms can realise their potential benefits but also mitigate their risks, as well as the tools necessary to make them more widely accepted including tools to address bias and potential accountability and transparency measures (as we discussed in Chapters 2 and 3).***

70. Our inquiry has also identified other key areas which, we believe, should be prominent in the Centre's early work. It should, as we described in Chapter 2, examine the biases built into algorithms—to identify, for example, how better 'training data' can be used; how unjustified correlations are avoided when more meaningful causal relationships are discernible; and how algorithm developer teams should be established which include a sufficiently wide cross-section of society, or of the groups that might be affected by an algorithm. The new body should also, we recommend, evaluate accountability tools—principles and 'codes', audits of algorithms, certification of algorithm developers, and charging ethics boards with oversight of algorithmic decisions—and advise on how they should be embedded in the private sector as well as in government bodies that share their data with private sector developers (Chapter 3).

235 The Royal Society (ADM0021)
236 University College London (ALG0050)
237 Future Advocacy (ALG0064)
238 TechUK (ADM0003)
239 HM Government, '*Tech sector backs British AI industry with multi million pound investment*', 26 April 2018

71. There are also important and urgent tasks that the Centre for Data Ethics & Innovation should address around the regulatory environment for algorithms; work which requires priority because of the Cambridge Analytica case, uncertainty about how the General Data Protection Regulation (GDPR) will address the issues around the use of algorithms, and because of the widespread and rapidly growing application of algorithms across the economy.

72. Cambridge Analytica allegedly harvested personal data from Facebook accounts without consent.[240] Through a personality quiz app, set up by an academic at the University of Cambridge, 270,000 Facebook users purportedly gave their consent to their data being used. However, the app also took the personal data of those users' 'friends' and contacts—in total at least 87m individuals. It has been reported that firms linked to Cambridge Analytica used these data to target campaign messages and sought to influence voters in the 2016 EU Referendum, as well as elections in the US and elsewhere.[241] The Information Commissioner[242] and the Electoral Commission[243] have been investigating the Cambridge Analytica case.

'Automated' decisions

73. The GDPR will have a bearing on the way algorithms are developed and used, because they involve the processing of data. Article 22 of the GDPR prohibits many uses of data processing (including for algorithms) where that processing is 'automated' and the 'data subject' objects. It stipulates that:

> The data subject shall have the right not to be subject to a decision based solely on automated processing, including 'profiling', which produces legal effects concerning him or her or similarly significantly affects him or her.[244]

The ICO explained that "unless it's (i) a trivial decision, (ii) necessary for a contract or (iii) authorised by law, organisations will need to obtain explicit consent to be able to use algorithms in decision-making". They believed that the GDPR provides "a powerful right which gives people greater control over automated decisions made about them".[245] The minister saw this as a positive step, explaining that:

> People must be informed if decisions are going to be made by algorithms rather than human management. Companies must make them aware of that.[246]

The Data Protection Bill provides a right to be informed, requiring data controllers to "notify the data subject in writing that a [significant] decision has been taken based solely on automated processing". This is to be done "as soon as reasonably practicable". If the data subject then exercises their right to opt-out, the Bill also allows the individual to request

240 '*FTC to question Facebook over Cambridge Analytica data scandal*', FT, 20 March 2018; New York Times, '*Facebook's Surveillance Machine*', 19 March 2018

241 '*The great British Brexit robbery: how our democracy was hijacked*', The Guardian, 7 May 2017

242 ICO, *The Information Commissioner opens a formal investigation into the use of data analytics for political purposes*' May 2017; ICO, *Update on ICO investigation into data analytics for political purposes*, Dec 2017.

243 Electoral Commission, *FOI release*, May 2017

244 (EU) 2016/679, Article 22

245 Information Commissioner's Office (ALG0038)

246 Q371

either that the decision be reconsidered or that a "new decision that is not based solely on automated processing" is considered. However, this is limited to decisions 'required or authorised in law' and would be unavailable for the vast majority of decisions."

74. Dr Sandra Wachter of the Oxford Internet Institute told us that what constituted the term 'significant affect' in the GDPR was "a very complicated and unanswered question".[247] Guidance from the relevant GDPR working party, an independent European advisory body on data protection and privacy, explained that:

> The decision must have the potential to significantly influence the circumstances, behaviour or choices of the individuals concerned. At its most extreme, the decision may lead to the exclusion or discrimination of individuals.[248]

Ultimately, Dr Wachter told us, "it will depend on the individual circumstances of the individual".[249]

75. Silkie Carlo, then from Liberty, had concerns about the law-enforcement derogations, which she believed should not apply to decisions affecting human rights: "The GDPR allows member states to draw their own exemption. Our exemptions have been applied in a very broad way for law enforcement processing and intelligence service processing in particular. That is concerning."[250] Others have criticised the fact that it is the data subject themselves that will have to "discern and assess the potential negative outcomes of an automated decision" when the "algorithms underlying these decisions are often complex and operate on a random-group level".[251]

76. The restriction of Article 22 of the GDPR to decisions 'based solely on automated processing' concerned the Institute of Mathematics and its Applications. They highlighted that many algorithms may "in principle only be advisory", and therefore not 'automated', "but the human beings using it may in practice just rubber-stamp its 'advice', so in practice it's determinative".[252] University College London was similarly concerned that decisions may be effectively 'automated' because of "human over-reliance on machines or the perception of them as objective and/or neutral", while the protections of Article 22 would "fall away".[253] Professor Kate Bowers the UCL Jill Dando Institute worried similarly that "people could just pay lip service to the fact that there is a human decision" involved in algorithmic processes.[254]

77. The GDPR working party on Article 22 recommended that "unless there is 'meaningful human input', a decision should still be considered 'solely' automated. This requires having individuals in-the-loop who a) regularly change decisions; and b) have the authority and competence organisationally to do so without being penalised."[255] Durham Constabulary told us that its HART algorithm (paragraph 21) only "supports decision-

247 Q62
248 Article 29 Data Protection Working Party, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (October 2017), p 10
249 Q62
250 Q52
251 University of Leuven Centre for IT and IP, *The Right not to be Subject to Automated Decision-Making: The role of explicit consent*, 2 August 2016
252 Institute of Mathematics and its Applications (ADM0008) para 8
253 University College London (ADM0010) para 5
254 Q175
255 University College London (ADM0010) para 6

making for the custody officer"[256] and that a human always remains in the loop. It was running a test of the algorithm's reliability by comparing its results against police officers making unaided decisions in parallel.[257]

78. The sort of algorithm used in the Cambridge Analytica case would be effectively prohibited when the GDPR's 'automated' processing provisions become effective in May 2018 if, as has been reported, the algorithm was used to target political campaign messages without human intervention.

Consent

79. Even if future use of the Cambridge Analytica algorithm would not be regarded as 'automated', and therefore a potentially allowable use of data, it would have to satisfy the requirements of the GDPR on consent.

80. The GDPR seeks to embed 'privacy by design' by addressing data protection when designing new data-use systems.[258] The ICO told us that "in data protection terms, transparency means that people should be given some basic information about the use of their personal data, such as the purpose of its use and the identity of the organisation using it."[259] The GDPR addresses online 'terms and conditions' clauses which are often used to get consent. As our predecessor Committee explained, the way these are used has significant shortcomings.[260] In our current inquiry too, Dr Sandra Wachter of the Oxford Internet Institute pointed out that few people would go through "hundreds of pages" of terms and conditions, and she instead preferred to see an "understandable overview of what is going to happen to your data while you are visiting a service".[261] The Minister, Margot James, also acknowledged the importance of "active consent", and emphasised the introduction of opt-outs in the GDPR as a mechanism for achieving this.[262] Our predecessor Committee highlighted the potential of "simple and layered privacy notices to empower the consumer to decide exactly how far they are willing to trust each data-holder they engage with".[263] In our inquiry, Dr Pavel Klimov suggested that such 'layered notices' could be helpful, giving certain critical information up-front and then allowing the user to click further if they want to learn more, including policies on sharing data with third-parties.[264]

81. Algorithm technology might in the future be used itself to provide transparency and consent by notifying data subjects when their data are used in other algorithms. DeepMind told us that they were working on a 'verifiable data audit' project using digital ledgers ('blockchains') to give people cryptographic proof that their data are being used in particular ways.[265]

256 Q150
257 Sheena Urwin, Head of Criminal Justice, Durham Constabulary (ADM0032)
258 EU GDPR, '*GDPR Key Changes*', accessed 20 March 2018
259 Information Commissioner's Office (ALG0038)
260 Science and Technology Committee, Fourth Report of Session 2015–16, *The big data dilemma*, HC 468
261 Q57
262 Q365
263 Science and Technology Committee, Fourth Report of Session 2015–16, *The big data dilemma*, HC 468, para 66
264 Q66
265 Q245 [Dr Dominic King]

82. In the meantime, privacy and consent remain critical issues for algorithms—just as they are (as our predecessor Committee found) for compiling profiles of people from diverse 'big data' datasets—because personal data are not always sufficiently anonymised. As our previous Committee highlighted, the risk on 'big data' analytics has been that data anonymisation can be undone as datasets are brought together.[266] Such risks apply equally to algorithms that look for patterns across datasets, although Dr M-H. Carolyn Nguyen of Microsoft argued that anonymisation could still play a part in deterring privacy abuse provided it is backed up by privacy laws.[267]

83. Cambridge Analytica's use of personal data, if used in the UK as has been alleged, would not have met the requirements for consent, even under the existing (pre-GDPR) regime. While it harvested the personal data of at least 87 million users, only the 270,000 individuals who were participants in the initial 'personality survey' were asked for consent.[268] The provisions of the GDPR will be applied where the 'data processor' or the data processing itself is in EU countries (or in the UK through the Data Protection Bill), or if individuals ('data subjects') are in the EU/UK.

84. In situations where consent is obtained, there is problem of the power imbalance between the individual and the organisation seeking consent. According to the Information Commissioner, "we are so invested" in digital services that "we become dependent on a service that we can't always extricate ourselves from".[269] This is especially true where, through acquisitions, companies restrict alternative services, as the Information Commissioner goes on to say. The £11 million of research announced in the AI sector deal (paragraph 6) is intended to better understand the "ethical […] implications of data sharing".[270]

Data protection impact assessments

85. To help identify bias in data-driven decisions, which we examined in Chapter 2, the GDPR requires 'data protection impact assessments'. Article 35 of the GDPR, reflected in the Data protection Bill, states:

> Where a type of [data] processing […] is likely to result in a high risk to the rights and freedoms of natural persons, the [data] controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data.[271]

Elizabeth Denham, the Information Commissioner, expected such impact assessments to be produced "when they are building AI or other technological systems that could have an impact on individuals".[272] According to the GDPR working party, these impact assessments offer "a process for building and demonstrating compliance",[273] and the Information Commissioner hoped that they would "force the organisation to think

266 *The big data dilemma*, HC 468; The Royal Society (ALG0056)
267 Q139
268 '*FTC to question Facebook over Cambridge Analytica data scandal*', Financial Times, 20 March 2018
269 '*Why we should worry about WhatsApp accessing our personal information*', The Guardian, 10 November 2016
270 HM Government, '*Tech sector backs British AI industry with multi million pound investment*', 26 April 2018
271 (EU) 2016/679, Article 35
272 Q303
273 Article 29 Data Protection Working Party, Processing is "likely to result in a high risk" for the purposes of Regulation 2016/679 (April 2017), p 4

through carefully what data are going into an AI system, how decisions are going to be made and what the output is."[274] Because of "commercial sensitives", however, she would not be "promoting the need to publish" the assessments.[275]

86. It is arguable whether those Facebook users who completed the personality questionnaire, that Cambridge Analytica subsequently used to target campaigning material, gave their full, informed consent. It is clear, however, that the millions of people receiving material because their data were included in the algorithm as 'friends' or contacts of those completing the questionnaire did not give their consent. Recently, in dealing with the Cambridge Analytica controversy, Facebook has begun to provide its customers with an explicit opportunity to allow or disallow apps that use their data. Every 90 days, users will be prompted to a Facebook Login process where users can specify their data permissions.[276] Whether Facebook or Cambridge Analytica would have undertaken a 'data protection impact assessment' to meet the requirements of the GDPR is impossible to know. It appears to us, however, that had they completed such an assessment they would have concluded that the algorithm would have been 'likely to result in a high risk to the rights and freedoms' of the individuals affected.

The Information Commissioner's powers

87. Enforcement of key features of the GDPR will fall on the shoulders of the Information Commissioner.[277] Professor Louise Amoore of Durham University expressed misgivings that "the ICO was only able to ask questions about how the data was being used and not the form of analysis that was taking place".[278] In a December 2017 speech, however, the Information Commissioner said that the ICO's duties were "wide and comprehensive, and not merely a complaints-based regulator. […] My office is here to ensure fairness, transparency and accountability in the use of personal data on behalf of people in the UK".[279] The GDPR will provide the Information Commissioner with greater powers, including under Article 58 to undertake data protection audits, as well as the right to obtain all personal data necessary for the ICO's investigations and to secure access to any premises required.[280] The GDPR will also give the ICO the power to ban data processing operations, and to issue much more significant financial penalties than under the existing regulations.[281]

88. Under the GDPR, however, the ICO cannot compel companies to make their data available. In March 2018 the Information Commissioner issued a "Demand for Access to records and data in the hands of Cambridge Analytica", but had to secure a High Court warrant to gain access to the data when the company did not oblige.[282] The delay in the ICO's access led some to question the powers of the Information Commissioner to quickly obtain 'digital search warrants'.[283] In her submission to the Data Protection Bill Committee in March 2018, the Information Commissioner wrote:

274 Q303
275 Q304
276 Facebook - Developer News, '*User Access Token Changes*', 09 April 2018
277 Information Commissioner's Office (ALG0038)
278 Q27
279 *Speech by the Information Commissioner at the TechUK Data Ethics Summit*, 13 December 2017
280 (EU) 2016/679, Article 58
281 (EU) 2016/679, Article 83; Information Commissioner's Office (ALG0038)
282 '*ICO statement: investigation into data analytics for political purposes*'.Accessed: 24 March 2018
283 Financial Times, '*UK data watchdog still seeking Cambridge Analytica warrant*', 20 March 2018

> Under the current Data Protection Act (DPA 1998), non-compliance with an Information Notice is a criminal offence, punishable by a fine in the Magistrate's Court. However, the court cannot compel compliance with the Information Notice or issue a disclosure order. This means, that although the data controller can receive a criminal sanction for non-compliance, the Commissioner is still unable to obtain the information she needs for her investigation.[284]

She complained that the inability to compel compliance with an Information Notice meant that investigations have "no guarantee of success", and which "may affect outcomes as it proves impossible to follow essential lines of enquiry". She contrasted this with her previous role as the Information and Privacy Commissioner for British Columbia where she had a power "to compel the disclosure of documents, records and testimony from data controllers and individuals, and failure to do so was a contempt of court". As a result, she called for the Data Protection Bill to "provide a mechanism to require the disclosure of requested information under her Information Notice powers". In her opinion, "Failure to do this will have an adverse effect on her investigatory and enforcement powers."[285] Addressing these challenges, the Government subsequently amended the Bill to increase the Information Commissioner's powers; enabling the courts to compel compliance with Information Orders and making it an offence to "block" or otherwise withhold the required information.[286]

89. The developments within algorithms and the way data are used have changed since the Information Commissioner's Office was set up. To accommodate this new landscape, Hetan Shah called "for Government to sort out its funding model".[287] The Government has since "announced a new charging structure" requiring large organisation to pay a higher fee, representative of their higher risk.[288]

90. **The provisions of the General Data Protection Regulation will provide helpful protections for those affected by algorithms and those whose data are subsumed in algorithm development, although how effective those safeguards are in practice will have to be tested when they become operational later this spring. While there is, for example, some uncertainty about how some of its provisions will be interpreted, they do appear to offer important tools for regulators to insist on meaningful privacy protections and more explicit consent. The Regulation provides an opt-out for most 'automated' algorithm decisions, but there is a grey area that may leave individuals unprotected—where decisions might be indicated by an algorithm but are only superficially reviewed or adjusted by a 'human in the loop', particularly where that human intervention is little more than rubber-stamping the algorithms' decision. While we welcome the inclusion in the Data Protection Bill of the requirement for data controllers to inform individuals when an automated algorithm produces a decision, it is unfortunate that it is restricted to decisions 'required or authorised by law'. There**

284 '*Information Commissioner's Office (DPB05)*', Data Protection Bill, Public Bill Committee, March 2018
285 '*Information Commissioner's Office (DPB05)*', Data Protection Bill, Public Bill Committee, March 2018
286 '*HL Bill 104 Commons amendments*', 09 May 2018, p 6,7
287 Q44
288 Information Commissioner's Office, '*New model announced for funding the data protection work of the Information Commissioner's Office*', 21 February 2018

is also a difficulty in individuals exercising their right to opt-out of such decisions if they are unaware that they have been the subject of an entirely automated process in the first place.

91. ***The Centre for Data Ethics & Innovation and the ICO should keep the operation of the GDPR under review as far as it governs algorithms, and report to Government by May 2019 on areas where the UK's data protection legislation might need further refinement. They should start with a more immediate review of the lessons of the Cambridge Analytica case. We welcome the amendments made to the Data Protection Bill which give the ICO the powers it sought in relation to its Information Notices, avoiding the delays it experienced in investigating the Cambridge Analytica case. The Government should also ensure that the ICO is adequately funded to carry out these new powers. The Government, along with the ICO and the Centre for Data Ethics & Innovation, should continue to monitor how terms and conditions rules under the GDPR are being applied to ensure that personal data is protected and that consumers are effectively informed, acknowledging that it is predominantly algorithms that use those data.***

92. **'Data protection impact assessments', required under the GDPR, will be an essential safeguard.** ***The ICO and the Centre for Data Ethics & Innovation should encourage the publication of the assessments (in summary form if needed to avoid any commercial confidentiality issues). They should also consider whether the legislation provides sufficient powers to compel data controllers to prepare impact assessments, and to improve them if the ICO and the Centre believe the assessments to be inadequate.***

Sector regulation

93. There is a wider issue for the Centre for Data Ethics & Innovation to consider early in its work, we believe, about any role it might have in providing regulatory oversight to complement the ICO's remit.

94. Nesta advocated the establishment of "some general principles around accountability, visibility and control" but applied with "plenty of flexibility". They believed that it was now time "to start designing new institutions".[289] The Financial Services Consumer Panel also wanted "a framework in place for supervision and enforcement as algorithmic decision making continues to play an increasing role in the financial services sector".[290] The Royal Society concluded that: "The volumes, portability, nature, and new uses of data in a digital world raise many challenges for which existing data access frameworks do not seem well equipped. It is timely to consider how best to address these novel questions via a new framework for data governance."[291]

289 Nesta (ALG0059)
290 Financial Services Consumer Panel (ALG0039)
291 The Royal Society (ALG0056)

95. There was a range of views in our inquiry on the relative benefits of a general overarching oversight framework and a sector-specific framework. Nesta doubted the effectiveness of "well intentioned private initiatives" which would be "unlikely to have the clout or credibility to deal with the more serious potential problems".[292] The Royal Society favoured sectoral regulation:

> While there may be specific questions about the use of machine learning in specific circumstances, these should be handled in a sector-specific way, rather than via an overarching framework for all uses of machine learning.[293]

They noted that the impact of algorithms which affect "buying or listening recommendations" matter less than those filtering what "appears to me as news, or affect how I am evaluated by others".[294] Similarly, Professor Kate Bowers of the UCL Jill Dando Institute believed that algorithms are context specific and that there is "a different set of risks and issues from the point of view of the degree to which they expose individuals".[295] These arguments suggest sectoral regulation as opposed to having a single regulator—a view supported by Elizabeth Denham, the Information Commissioner, who did not think that we need "an AI regulator",[296] but was nevertheless bringing sector regulators together "to talk about AI systems".[297] This is a role that could be taken by the newly created Centre for Data Ethics and Innovation, a view also shared by the Minister, Margot James.[298]

96. In contrast to this sectoral approach, the Oxford Internet Institute proposed "an AI Watchdog, or a trusted and independent regulatory body" which would be "equipped with the proper expertise (spanning ideally law, ethics, to computer science), resources and auditing authority (to make inspections) to ensure that algorithmic decision making is fair, unbiased and transparent".[299] In a similar vein, Microsoft favoured "all aspects of society, including government, academia and business [… coming] together to create a set of shared principles by which to guide the use of algorithms and AI",[300] although not necessary leading to overarching regulation. Nesta wanted an advisory body to "guide behaviours, understanding, norms and rules", without "formal regulatory powers of approval or certification" but instead "strong powers of investigation and of recommendation".[301]

97. ***The Centre for Data Ethics & Innovation and the Information Commissioner should review the extent of algorithm oversight by each of the main sector-specific regulators, and use the results to guide those regulators to extend their work in this area as appropriate. The Information Commissioner should also make an assessment, on the back of that work, of whether it needs greater powers to perform its regulatory oversight role where sector regulators do not see this as a priority.***

292 Nesta (ALG0059)
293 The Royal Society (ADM0021)
294 Mark Gardiner (ALG0068). See also "*Should algorithms be regulated?*", IT Pro, 19 December 2016
295 Q156
296 Q298
297 Q300
298 Q378
299 Oxford Internet Institute (ALG0031)
300 Microsoft (ALG0072) para 11
301 Nesta (ALG0059)

Conclusions and recommendations

Introduction

1. The Government's proposed Centre for Data Ethics & Innovation is a welcome initiative. It will occupy a critically important position, alongside the Information Commissioner's Office, in overseeing the future development of algorithms and the 'decisions' they make. The challenge will be to secure a framework which facilitates and encourages innovation but which also maintains vital public trust and confidence. (Paragraph 7)

2. Many of the issues raised in this report will require close monitoring, to ensure that the oversight of machine learning-driven algorithms continues to strike an appropriate and safe balance between recognising the benefits (for healthcare and other public services, for example, and for innovation in the private sector) and the risks (for privacy and consent, data security and any unacceptable impacts on individuals). As we discuss in this report, the Government should ensure that these issues are at the top of the new body's remit and agenda. (Paragraph 8)

3. The Government plans to put the Centre for Data Ethics & Innovation on a statutory footing. When it does so, it should set it a requirement to report annually to Parliament on the results of its work, to allow us and others to scrutinise its effectiveness. Although the terms of the Government's proposed consultation on the Centre for Data Ethics & Innovation have yet to be announced, we anticipate our report feeding into that exercise. (Paragraph 9)

Applications and bias

4. Algorithms are being used in an ever-growing number of areas, in ever-increasing ways. They are bringing big changes in their wake; from better medical diagnoses to driverless cars, and within central government where there are opportunities to make public services more effective and achieve long-term cost savings. They are also moving into areas where the benefits to those applying them may not be matched by the benefits to those subject to their 'decisions'—in some aspects of the criminal justice system, for example, and algorithms using social media datasets. Algorithms, like 'big data' analytics, need data to be shared across previously unconnected areas, to find new patterns and new insights. (Paragraph 29)

5. *The Government should play its part in the algorithms revolution in two ways. It should continue to make public sector datasets available, not just for 'big data' developers but also algorithm developers. We welcome the Government's proposals for a 'data trusts' approach to mirror its existing 'open data' initiatives. Secondly, the Government should produce, publish, and maintain a list of where algorithms with significant impacts are being used within Central Government, along with projects underway or planned for public service algorithms, to aid not just private sector involvement but also transparency. The Government should identify a ministerial champion to provide government-wide oversight of such algorithms, where they are used by the public sector, and to co-ordinate departments' approaches to the development and deployment of algorithms and partnerships with the private sector.* (Paragraph 30)

6. Algorithms need data, and their effectiveness and value tends to increase as more data are used and as more datasets are brought together. The Government could do more to realise some of the great value that is tied up in its databases, including in the NHS, and negotiate for the improved public service delivery it seeks from the arrangements and for transparency, and not simply accept what the developers offer in return for data access. *The Crown Commercial Service should commission a review, from the Alan Turing Institute or other expert bodies, to set out a procurement model for algorithms developed with private sector partners which fully realises the value for the public sector. The Government should explore how the proposed 'data trusts' could be fully developed as a forum for striking such algorithm partnering deals. These are urgent requirements because partnership deals are already being struck without the benefit of comprehensive national guidance for this evolving field.* (Paragraph 31)

7. Algorithms, in looking for and exploiting data patterns, can sometimes produce flawed or biased 'decisions'—just as human decision-making is often an inexact endeavour. As a result, the algorithmic decision may disproportionately discriminate against certain groups, and are as unacceptable as any existing 'human' discrimination. Algorithms, like humans, can produce bias in their results, even if unintentional. When algorithms involve machine learning, they 'learn' the patterns from 'training data' which may be incomplete or unrepresentative of those who may be subsequently affected by the resulting algorithm. That can result, for example, in race or gender discrimination in recruitment processes. The patterns that algorithms rely on may be good correlations but may not in fact show a reliable causal relationship, and that can have important consequences if people are discriminated against as a result (such as in offender rehabilitation decisions). Algorithms may have incomplete data so that, for example, some do not get favourable financial credit decisions. Algorithm developer teams may not include a sufficiently wide cross-section of society (or the groups that might be affected by an algorithm) to ensure a wide range of perspectives is subsumed in their work. These biases need to be tackled by the industries involved and .... by the regulatory environment being introduced by the GDPR, and safeguards against bias should be a critical element of the remit of the Centre for Data Ethics & Innovation. (Paragraph 44)

Accountability and transparency

8. Setting principles and 'codes', establishing audits of algorithms, introducing certification of algorithms, and charging ethics boards with oversight of algorithmic decisions, should all play their part in identifying and tackling bias in algorithms. With the growing proliferation of algorithms, such initiatives are urgently needed. *The Government should immediately task the Centre for Data Ethics & Innovation to evaluate these various tools and advise on which to prioritise and on how they should be embedded in the private sector as well as in government bodies that share their data with private sector developers. Given the international nature of digital innovation, the Centre should also engage with other like-minded organisations in other comparable jurisdictions in order to develop and share best practice.* (Paragraph 56)

9. Transparency must be a key underpinning for algorithm accountability. There is a debate about whether that transparency should involve sharing the workings of the algorithm 'black box' with those affected by the algorithm and the individuals

whose data have been used, or whether (because such information will not be widely understood) an 'explanation' is provided. *Where disclosure of the inner workings of privately-developed public-service algorithms would present their developers with commercial or personal-data confidentiality issues, the Government and the Centre for Data Ethics & Innovation should explore with the industries involved the scope for using the proposed 'data trust' model to make that data available in suitably de-sensitised format. While we acknowledge the practical difficulties with sharing an 'explanation' in an understandable form, the Government's default position should be that explanations of the way algorithms work should be published when the algorithms in question affect the rights and liberties of individuals. That will make it easier for the decisions produced by algorithms also to be explained. The Centre for Data Ethics & Innovation should examine how explanations for how algorithms work can be required to be of sufficient quality to allow a reasonable person to be able to challenge the 'decision' of the algorithm. Where algorithms might significantly adversely affect the public or their rights, we believe that the answer is a combination of explanation and as much transparency as possible.* (Paragraph 66)

10. *The 'right to explanation' is a key part of achieving accountability. We note that the Government has not gone beyond the GDPR's non-binding provisions and that individuals are not currently able to formally challenge the results of all algorithm decisions or where appropriate to seek redress for the impacts of such decisions. The scope for such safeguards should be considered by the Centre for Data Ethics & Innovation and the ICO in the review of the operation of the GDPR that we advocate.* (Paragraph 67)

The Centre for Data Ethics & Innovation, research and the regulatory environment

11. We welcome the announcement made in the AI Sector Deal to invest in research tackling the ethical implications around AI. *The Government should liaise with the Centre for Data Ethics & Innovation and with UK Research & Innovation, to encourage sufficient UKRI-funded research to be undertaken on how algorithms can realise their potential benefits but also mitigate their risks, as well as the tools necessary to make them more widely accepted including tools to address bias and potential accountability and transparency measures.* (Paragraph 69)

12. The provisions of the General Data Protection Regulation will provide helpful protections for those affected by algorithms and those whose data are subsumed in algorithm development, although how effective those safeguards are in practice will have to be tested when they become operational later this spring. While there is, for example, some uncertainty about how some of its provisions will be interpreted, they do appear to offer important tools for regulators to insist on meaningful privacy protections and more explicit consent. The Regulation provides an opt-out for most 'automated' algorithm decisions, but there is a grey area that may leave individuals unprotected—where decisions might be indicated by an algorithm but are only superficially reviewed or adjusted by a 'human in the loop', particularly where that human intervention is little more than rubber-stamping the algorithms' decision. While we welcome the inclusion in the Data Protection Bill of the requirement for data controllers to inform individuals when an automated algorithm produces a

decision, it is unfortunate that it is restricted to decisions 'required or authorised by law'. There is also a difficulty in individuals exercising their right to opt-out of such decisions if they are unaware that they have been the subject of an entirely automated process in the first place (Paragraph 90)

13. *The Centre for Data Ethics & Innovation and the ICO should keep the operation of the GDPR under review as far as it governs algorithms, and report to Government by May 2019 on areas where the UK's data protection legislation might need further refinement. They should start with a more immediate review of the lessons of the Cambridge Analytica case. We welcome the amendments made to the Data Protection Bill which give the ICO the powers it sought in relation to its Information Notices, avoiding the delays it experienced in investigating the Cambridge Analytica case. The Government should also ensure that the ICO is adequately funded to carry out these new powers. The Government, along with the ICO and the Centre for Data Ethics & Innovation, should continue to monitor how terms and conditions rules under the GDPR are being applied to ensure that personal data is protected and that consumers are effectively informed, acknowledging that it is predominantly algorithms that use those data* (Paragraph 91)

14. 'Data protection impact assessments', required under the GDPR, will be an essential safeguard. *The ICO and the Centre for Data Ethics & Innovation should encourage the publication of the assessments (in summary form if needed to avoid any commercial confidentiality issues). They should also consider whether the legislation provides sufficient powers to compel data controllers to prepare impact assessments, and to improve them if the ICO and the Centre believe the assessments to be inadequate.* (Paragraph 92)

15. *The Centre for Data Ethics & Innovation and the Information Commissioner should review the extent of algorithm oversight by each of the main sector-specific regulators, and use the results to guide those regulators to extend their work in this area as appropriate. The Information Commissioner should also make an assessment, on the back of that work, of whether it needs greater powers to perform its regulatory oversight role where sector regulators do not see this as a priority.* (Paragraph 97)

Formal minutes

Tuesday 15 May 2018

Members present:

Norman Lamb, in the Chair

Vicky Ford	Stephen Metcalfe
Bill Grant	Carol Monaghan
Darren Jones	Damien Moore
Liz Kendall	Neil O'Brien

Draft Report (*Algorithms in decision-making*), proposed by the Chair, brought up and read.

Ordered, That the draft Report be read a second time, paragraph by paragraph.

Paragraphs 1 to 97 read and agreed to.

Summary agreed to.

Resolved, That the Report be the Fourth Report of the Committee to the House.

Ordered, That the Chair make the Report to the House.

Ordered, That embargoed copies of the Report be made available, in accordance with the provisions of Standing Order No. 134.

[Adjourned till Tuesday 22 May at 9.00 am

Witnesses

The following witnesses gave evidence. Transcripts can be viewed on the inquiry publications page of the Committee's website.

Tuesday 14 November 2017	*Question numbers*
Hetan Shah, Executive Director, Royal Statistical Society; **Professor Nick Jennings**, Royal Academy of Engineering; **Dr Adrian Weller**, Turing Fellow, Alan Turing Institute; and **Professor Louise Amoore**, Durham University.	Q1–44
Silkie Carlo, Senior Advocacy Officer, Liberty; **Dr Sandra Wachter**, Lawyer and Researcher in Data Ethics, AI, and Robotics, Oxford Internet Institute; and **Dr Pavel Klimov**, Chair, Law Society's Technology and the Law Group.	Q45–88
Tuesday 12 December 2017	
Martin Wattenberg, Senior Staff Research Scientist, Google AI team; **Charles Butterworth**, Managing Director for UK, Europe, Middle East and Africa, Experian; **Dr M-H Carolyn Nguyen**, Director of Technology Policy, Microsoft; and **Nick Pickles**, Head of Public Policy UK and Israel, Twitter.	Q89–148
Sheena Urwin, Head of Criminal Justice, Durham Constabulary; and **Professor Kate Bowers**, Academic Director, UCL Jill Dando Institute.	Q149–207
Tuesday 16 January 2018	
Dr Dominic King, Senior Staff Research Scientist and Clinical Lead, DeepMind Health; **Dr Ian Hudson**, Chief Executive, Medicines and Healthcare products Regulatory Agency (MHRA); **Professor Harry Hemingway**, Farr Institute of Health Informatics Research; and **Eleonora Harwich**, Head of Digital and Tech Innovation, Reform.	Q208–293
Tuesday 23 January 2018	
Elizabeth Denham, Information Commissioner.	Q294–351
Margot James MP, Minister for Digital and the Creative Industries, Department for Digital, Culture, Media and Sport; **Oliver Buckley**, Deputy Director, Digital Charter and Data Ethics, Department for Digital, Culture, Media and Sport; and **Andrew Elliot**, Data Protection Bill Manager, Department for Digital, Culture, Media and Sport.	Q352–393

Published written evidence

The following written evidence was received and can be viewed on the inquiry publications page of the Committee's website.

ADM numbers are generated by the evidence processing system and so may not be complete

1 BCS, The Chartered Institute for IT (ADM0004)

2 Big Brother Watch (ADM0012)

3 CompTIA (ADM0022)

4 Datanut Sciences Consultancy (ADM0013)

5 Department for Digital, Culture, Media and Sport (ADM0015)

6 Durham Constabulary (ADM0032)

7 Filament (ADM0014)

8 Financial Conduct Authority (ADM0030)

9 Google (ADM0016)

10 Guardian News and Media (ADM0001)

11 IBM (ADM0017)

12 Information Commissioner's Office (ADM0031)

13 Institute of Mathematics and its Applications (ADM0008)

14 Inter-Disciplinary Ethics Applied (IDEA) Centre, University of Leeds (ADM0025)

15 Marion Oswald (ADM0002)

16 medConfidential (ADM0007)

17 Medicines and Healthcare products Regulatory Agency (MHRA) (ADM0026)

18 Mozilla (ADM0005)

19 Mr John Phillips (ADM0024)

20 Mrs Katherine Garzonis (ADM0019)

21 Ofcom (ADM0028)

22 Ofgem (ADM0027)

23 Ofsted (ADM0029)

24 PHG Foundation (ADM0011)

25 Professor Ashiq Anjum (ADM0009)

26 REACT / REFLECT research team, University of Manchester (ADM0023)

27 techUK (ADM0003)

28 The Operational Research Society (ADM0006)

29 The Royal Society (ADM0021)

30 University College London (ADM0010)

31 University of Nottingham (ADM0018)

The following written evidence was received in the last Parliament by the previous Committee for this inquiry and can be viewed on the inquiry publications page of the Committee's website.

ALG numbers are generated by the evidence processing system and so may not be complete

1 Academy of Medical Sciences (ALG0055)

2 Aire (ALG0066)

3 BCS, The Chartered Institute for IT (ALG0053)

4 Centre for Intelligent Sensing (ALG0036)

5 Council of Professors and Heads of Computing (ALG0061)

6 DataKind UK (ALG0037)

7 Department for Culture, Media and Sport (ALG0047)

8 Dr Alex Griffiths, Dr Henry Rothstein and Prof David Demeritt (ALG0065)

9 Dr Alison Powell (ALG0067)

10 Dr Galina Andreeva and Dr hab. Anna Matuszyk (ALG0062)

11 Dr Janet Bastiman (ALG0029)

12 Dr Stephen Castell (ALG0052)

13 Durham Constabulary (ALG0041)

14 Financial Services Consumer Panel (ALG0039)

15 Future Advocacy (ALG0064)

16 Imosphere (ALG0044)

17 Information Commissioner's Office (ALG0038)

18 Institute of Mathematics and its Applications (ALG0028)

19 Liberty (ALG0070)

20 Marion Oswald and Sheena Urwin (ALG0030)

21 Microsoft (ALG0072)

22 Mr David Strudwick (ALG0040)

23 Mr Jamie Grace (ALG0003)

24 Mr Mark Gardiner (ALG0068)

25 Mr Tom Macfarlane (ALG0069)

26 Nesta (ALG0059)

27 Office for National Statistics (ALG0060)

28 Oxford Internet Institute, University of Oxford (ALG0031)

29 Paul Pedley (ALG0009)

30 Polygeia (ALG0043)

31 Professor Daniel Neyland (ALG0027)

32 Professor Louise Amoore (ALG0042)

33 Professor Nigel Harvey (ALG0054)

34 Projects by IF (ALG0033)

35 Research Councils UK (RCUK) (ALG0074)
36 Royal Academy of Engineering (ALG0046)
37 Sensible Code Company (ALG0010)
38 Simul Systems Ltd (ALG0007)
39 Space for Sharing Project (ALG0006)
40 The Alan Turing Institute (ALG0073)
41 The Human Rights, Big Data and Technology Project (ALG0063)
42 The Operational Research Society (ALG0045)
43 The Royal Society (ALG0056)
44 The Royal Statistical Society (ALG0071)
45 UCL Jill Dando Institute of Security and Crime Science (ALG0048)
46 University College London (ALG0050)
47 University of Nottingham and University of Oxford (ALG0049)

List of Reports from the Committee during the current Parliament

All publications from the Committee are available on the publications page of the Committee's website. The reference number of the Government's response to each Report is printed in brackets after the HC printing number.

Session 2017–19

First Report	Pre-appointment hearing: chair of UK Research & Innovation and executive chair of the Medical Research Council	HC 747
Second Report	Brexit, science and innovation	HC 705
Third Report	Genomics and and genome editing in the NHS	HC 349
First Special Report	Science communication and engagement: Government Response to the Committee's Eleventh Report of Session 2016–17	HC 319
Second Special Report	Managing intellectual property and technology transfer: Government Response to the Committee's Tenth Report of Session 2016–17	HC 318
Third Special Report	Industrial Strategy: science and STEM skills: Government Response to the Committee's Thirteenth Report of Session 2016–17	HC 335
Fourth Special Report	Science in emergencies: chemical, biological, radiological or nuclear incidents: Government Response to the Committee's Twelfth Report of Session 2016–17	HC 561
Fifth Special Report	Brexit, science and innovation: Government Response to the Committee's Second Report	HC 1008