Artificial Intelligence and Public Standards

Written evidence

Submission	Name
1	Christopher Marsh, Bale Crocker
2	Jamie Grace, Sheffield Hallam University
3	Simon Dennis, SAS Institute
4	Marion Oswald, Northumbria University
5	MedConfidential
6	Crofton Black, Bureau of Investigative Journalism
7	DittoAI
8	Royal College of Physicians
9	Leverhulme Centre for the Future of Intelligence, University of Cambridge
10	David Evans, Goodfaih
11	British Computer Society
12	Dr Emma Carmel, University of Bath
13	Robin Allen QC and Dee Masters, Cloisters Chambers
14	CarnegieUK Trust
15	Centre for Information Policy Leadership
16	Chartered Institute of Public Relations
17	Mission and Public Affairs Council of the Church of England
18	Centre for Data Ethics and Innovation
19	Information Commissioner's Office
20	Professor Karen Yeung, University of Birmingham

SUBMISSION 1

Dear Sir/Madam

(A) Regarding your notice of 25 March 2019 from the Committee on Standards in Public Life entitled *Review of artificial intelligence and public standards* I submit for your consideration my 2016 paper *Integrity* at www.balecrocker.com/integrity subtitled *What can trusted advisers offer that AIs can't?* In my view now, the premise that human integrity is by its nature not surpassable by that of artificial intelligences is at best uncertain, but the argument and sources might have interest.

(B) Furthermore, I submit the view that human intellectual endeavour must ride the wave of AIs just as it has done so in the case of other tools. We *must expect* to develop mental capabilities that ensure a proper relationship to AI tools. We *must expect* government to facilitate and lead in this precise field as a matter of urgency. I put this widely held case in an internal talk to the Bodleian Library on 29 March 2019, and also in one of the first editions of the journal of the British Interactive Media Association 30 years ago.

(C) Furthermore, I submit the view that AIs directly impact the seven principles and their interpretation. Comments follow.

> 1. Selflessness
> Holders of public office should act solely in terms of the public interest.
> Comment: The public interest is for a body of people in public service to maintain intellectual superiority over machine intelligence.
>
> 2. Integrity
> Holders of public office must avoid placing themselves under any obligation to people or organisations that might try inappropriately to influence them in their work.
> Comment: Depending on the interpretation of "obligation" is it doubtful whether such holders can comply now with that requirement if they use existing digital tools that depend on current 'AIs'?
>
> They should not act or take decisions in order to gain financial or other material benefits for themselves, their family, or their friends.
> Comment: No problem there, although the impact of online advertising practice on the interpretation of "in order to" must be carefully watched.
>
> They must declare and resolve any interests and relationships.
> Comment: Personal data trails must be taken into account here as a form of "relationship(s)" with AIs and their owners and their operators and those with commercial or other engagement with them.
>
> 3. Objectivity
> Holders of public office must act and take decisions impartially, fairly and on merit, using the best evidence and without discrimination or bias.

Comment: The entirely appropriate and no doubt very well understood, respected and managed burden of this obligation must be increasing very rapidly as a result of increasing use of AIs in everyday tools. Such holders must be protected in this context.

4. Accountability
Holders of public office are accountable to the public for their decisions and actions and must submit themselves to the scrutiny necessary to ensure this.
Comment: (repeated) The entirely appropriate and no doubt very well understood, respected and managed burden of this obligation must be increasing very rapidly as a result of increasing use of AIs in everyday tools. Such holders must be protected in this context.

5. Openness
Holders of public office should act and take decisions in an open and transparent manner. Information should not be withheld from the public unless there are clear and lawful reasons for so doing.
Comment: (repeated) Comment: The entirely appropriate and no doubt very well understood, respected and managed burden of this obligation must be increasing very rapidly as a result of increasing use of AIs in everyday tools. Such holders must be protected in this context.

6. Honesty
Holders of public office should be truthful.
Comment: The time is right to clearly define the word *truthful* for such holders in our times.

7. Leadership
Holders of public office should exhibit these principles in their own behaviour. They should actively promote and robustly support the principles and be willing to challenge poor behaviour wherever it occurs.
Comment: Absolutely, and by refining this code they will become leaders and exemplars in pursuit of the mastery of machines that people must retain.

SUBMISSION 2

Written evidence submission to the Committee on Standards in Public Life inquiry into artificial intelligence and public standards

Jamie Grace, Sheffield Hallam University

April 17th 2019

Summary

1. Policymakers and political leaders in criminal justice settings, including Police and Crime Commissioners and their most senior police officer counterparts in forces themselves, must push each other to be more transparent over time, in relation to their openness as to the extent and the nature of the use of algorithmic tools in UK policing.

2. Policymakers must foster a climate of openness and transparency when it comes to the public protection priorities of setting the 'trade-offs' inherent in designing the differentiated accuracies of algorithmic intelligence analysis tools. This would better ensure that such tools are correctly aligned in terms of the public interest in their use, and the weighting that they demonstrate toward either statistical *specificity* or statistical *sensitivity*[1]. In essence we need transparency about the *type* of accuracy that a machine learning tool should seek to achieve.

3. Police organisations, for example, could use possible administrative law principles (concerning duties to give reasons, and to consult, and to consider and collect evidence of equality impacts as well as data protection impacts) to head off ruinous scrutiny from the courts. It may well be that by using those administrative law principles as a framework to make themselves more accountable, policing leaders can more rigorously adhere to the 'Nolan Principles'.

About my roles and research

I am currently a Senior Lecturer in Law in the Department of Law and Criminology at Sheffield Hallam University, holding this post since January 2014. I am course leader of both the MA and LL.M in Applied Human Rights courses taught in my Department. I am an active researcher in the Helena Kennedy Centre for International Justice, and a Fellow of the Sheffield Institute of Policy Studies, both part of Sheffield Hallam University. I have been appointed to the independent Data Analytics Ethics Committee established by the West Midlands Police and Crime Commissioner. I will hold a Visiting Fellowship at the Institute of Advanced Legal Studies (University of London) in the summer of 2020; as part of the Centre for Information Law and Policy at the IALS.

Some of my co-authored research on the legalities of the algorithmic analysis of police intelligence (Oswald et al, 2018) has been developed with the assistance of Durham Constabulary. It has been discussed with the Home Office, the Centre for Data Ethics and

[1] Trade-offs' are the notion used in the relevant academic literature to describe the pros-and-cons between building a predictive model which might look for one of two *types of accuracy*. This is because a key issue in the use of many potential machine learning deployments in the policing context is the notion that one type of accuracy, or the other, is to be given more weight: Statistical *specificity* (greater accuracy overall in correctly sorting high-, medium-, and low-risk offenders) or statistical *sensitivity* (detecting as many high-risk potential recidivists as possible, and tolerating a higher rate of 'false positives' as a result over time).

Innovation (DCMS), Her Majesty's Inspectorate of Fire and Rescue Services, the National Police Chiefs' Council, and the National Crime Agency; as well as the NGOs Big Brother Watch and Liberty, as well as forming the basis of conference papers presented at academic conferences at Oxford, Sheffield Hallam, Keele and Winchester Universities.

Key principles for algorithmic justice: Objectivity; Openness; Accountability

The key 'Nolan Principles' which my research work on algorithmic justice and the deployment of machine-learning tools in the law enforcement context has connected with most centrally are those values of: *Objectivity*; *Openness*; and *Accountability*. This written evidence has adapted the key points on these principles found in my body of written research, to date, on avoiding 'algorithmic impropriety'[2].

Objectivity

A crucial aspect of the use of algorithmic tools in criminal justice contexts is the extent to which safeguards exist for the maintenance of human decision-maker and policymaker oversight of the manner in which such tools are deployed. Drawing on our research that produced the ALGO-CARE regulatory framework tool in conjunction with Durham Constabulary, my research colleagues and I would ask of policymakers informed in their decision-making by an algorithmic tool the following key pair of questions concerning the need for a particular degree of human objectivity:

- Is the assessment made by the algorithm used in an advisory capacity, and therefore does a human officer retain decision-making discretion? (*Care should be taken to ensure that an algorithm is not inappropriately fettering an officer's discretion, as natural justice and procedural fairness claims may well arise.)*
- What other decision-making by human officers will add objectivity to the decisions (partly) based on the algorithm? (Police forces should *consider if supposedly advisory algorithmic assessments might in practice have an undue influence. If it is proposed that an algorithmic decision is automated and determinative then data protection rights in regard to automated decisions will apply.)*

For example, in terms of algorithmic policing tools being used to manage victim complainants and their reports of crime, the most high-profile example to date is the 'EBIT' system of case handling and investigation triage being used in trial areas by Kent Police, and which has been inspected favourably by HMICFRS, following only a little inquiry (based on a small number of interviews) into whether victims would object to an algorithm being involved in 'shelving' their report of a crime. The process and legal position of explaining to, or otherwise notifying victims that an algorithm has recommended their case or complaint

[2] J. Grace, 'Human rights, regulation and the right to restrictions on algorithmic police intelligence analysis tools in the UK', available online as a draft paper at: http://ssrn.com/abstract=3303313; J. Grace, 'Algorithmic impropriety in UK policing?', (2019) *Journal of Information Rights, Policy and Practice*; Vol, 3 Issue 1: https://jirpp.winchesteruniversitypress.org/articles/abstract/23/; and M. Oswald, J. Grace, S. Urwin & G. C. Barnes (2018) Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental' proportionality, *Information & Communications Technology Law*, Volume 27, 2018 - Issue 2, 223-250.

be shelved is a delicate and difficult one. At the time of writing a piece of research in 2018, I could find no clear information on the Kent Police website as to the EBIT system, and so it is unclear how individuals could at that time easily find out more about the role of the EBIT system in handling their complaint or report of a criminal offence.
It is likely that tools like EBIT in Kent be used with more serious offence types over time by that force and others. For serious sexual and violent offences however, one would imagine that ultimately there would be public policy barriers, and perhaps eventually, legal barriers to the regular and standardised use of an algorithmic tool which allows the police a *statistical* rationale to side-line and de-prioritise complaints of the worst crimes.
As things stand, the EBIT system currently, and sensibly, makes recommendations for a specialist team of officers in Kent to apply using their professional judgment, in terms of deciding whether or not to commit investigative resources to a report of an offence. Thus the Kent E-BIT system of oversight of victims' complaints and resulting investigations, resulting in human professional judgment being applied before an investigation into a serious offence is discontinued and a case 'shelved', is a form of 'human in the loop' regulation of the EBIT algorithm that simply has to be maintained as a core standard as a result.

Openness
Political leaders and policymakers in criminal justice settings, including Police and Crime Commissioners and their senior police officer counterparts in police service organisations, must push each other to be more transparent over time, in relation to their openness as to the extent and the nature of the use of algorithmic tools. A leading example that has emerged, at the time of writing, is the way that the West Midlands Police and Crime Commissioner and their advisors have created an Independent Data Analytics Ethics Committee. This Committee has already added an important layer of scrutiny to the proposal by West Midlands Police to deploy a particular 'Integrated Offender Management' tool.

More broadly, and outside of the UK, the regulatory picture in Europe is changing, too - with an increasing human rights focus on the issues of algorithmically-informed policing policy and deployments. The Council of Europe in November 2018 published draft *Guidelines for States on actions to be taken vis-à-vis the human rights impacts of algorithmic systems*, which could be taken into account by domestic courts in countries across Europe, including in the UK, when dealing with fresh cases on the regulation and impact of 'algorithmic police intelligence analysis tools' (APIATs) in their respective criminal justice systems. There is also an emphasis in these draft Guidelines on the need for member states within the Council of Europe to foster a climate of openness and transparency when it comes to the public protection priorities of setting the 'trade-offs' inherent in designing the differentiated accuracies of APIATs - and even the determination of which uses of such tools are not appropriate at all, or on balance[3]. The two chief recommendations found in the draft Guidelines with regard to these variations on the key Nolan Principle of transparency are:

> "**Recommendation 6.6…** States should engage in inclusive, inter-disciplinary, informed and public debates to define what areas of public services profoundly affecting access to or exercise of human

[3] n 1.

rights may not be appropriately determined, decided or optimised through algorithmic systems." (Council of Europe, 2018:9)

"**Recommendation 3.1…** States should thrive to ensure that all their data analysis and modelling of algorithmic systems that may profoundly affect the exercise of human rights is designed towards minimising negative impacts and maximising benefits for individuals and society. Evaluation before and after deployment should be part and parcel of these efforts and should include an evaluation of the desirability and legitimacy of the goal that the system intends to achieve or optimise." (Council of Europe, 2018:6)

Accountability

The bodies in the criminal justice or law enforcement context that are going to increasingly be using APIATs would be amenable to judicial review of the manner in which impactful decisions on individuals have been 'algorithmically informed', from the perspective of 'natural justice' or 'procedural fairness' arguments. In this way, administrative law arguments could readily be used to enhance and augment data protection and/or human rights grounds of judicial review in this emerging algorithmic era. At the moment, the lack of accountability over the use of algorithmic tools in criminal justice contexts is best represented by the Information Commissioner's Office issuing an Enforcement Notice in relation to the operation of the Gangs Matrix by the Metropolitan Police Service. The Mayor's Office of Policing and Crime (MOPAC, London) report released following the MPS Gangs Matrix Decision Notice from the ICO in November 2018 concluded that (p.55):

"Both the Operating Model and the training should have a particular focus on ensuring:

- "that the right people are on the Matrix;
- "that people are added and removed in a standardised, evidence-based manner;
- "that they can be removed and that the 'gang' label will not 'follow' them;
- "that local Matrices are refreshed regularly so that individuals don't stay on any longer than necessary…

Using judicial review, a campaigning organisation or an individual claimant could seek to obtain a remedy from the UK courts in relation to a certain biased or heavily skewed outcome of an algorithmically-based decision-making process or outcome in the criminal justice context. But police organisations, for example, could use possible administrative law principles (concerning duties to give reasons, and to consult, and to consider equality impacts as well as data protection impacts) to head off ruinous scrutiny from the courts by using those administrative law principles as a framework to make themselves more accountable.

Concluding points about ALGO-CARE

My co-authored evaluation of the legalities of the 'Harm Assessment Risk Tool' (HART), used currently by Durham Constabulary, led to the development of a model regulatory framework for algorithmic decision-making in criminal justice ('ALGO-CARE', see Appendix 1) which was discussed in a hearing of the House of Commons Science and Technology Committee in December 2017. My written evidence to this inquiry on algorithmic decision-making in the UK public sector was cited in the final inquiry report. 'ALGO-CARE' was also cited favourably in a report on artificial intelligence by the Lords Select Committee on Artificial Intelligence in April 2018. I was also invited to deliver a session on the future regulation of algorithmic analysis in criminal justice at a conference at the University of Oxford in November 2017, as well as at a roundtable in April 2018 organised by the Data

Justice Lab (University of Cardiff) in London, both attended by a wide range of public bodies and industry stakeholders. ALGO-CARE has been more recently been cited by a report for the Council of Europe on human rights standards for algorithms[4], and in reports by both the Police Foundation[5] and by the human rights group Liberty[6].
This dissemination of the research underpinning the ALGO-CARE framework has led to a degree of recognition and adoption of the model as a form of self-regulation across some parts of the police service:

- The National Police Chiefs' Council have now recommended to UK police forces that they adopt the ALGO-CARE model as an interim safeguard in determining whether and how to deploy AI in operational or strategic ways.
- Police organisations have begun to adopt 'ALGO-CARE' in their evaluation of machine-learning tools: West Midlands Police (policing a population of 2.8m people and home to the National Data Analytics Solution project); Durham Constabulary, as well as the National Police Chiefs' Council.
- In 2019 I have been organising a series of seminars and workshops with police forces to help them better understand issues in algorithmically-informed justice contexts, drawing on the 'ALGO-CARE' tool I helped to develop. This series of workshops, funded by Sheffield Hallam University, has seen me discuss and introduce the ALGO-CARE framework of self-regulation to the following forces and police organisations at the time of writing: the National Crime Agency; City of London Police, and constabularies from Avon and Somerset; Devon and Cornwall, Dorset, Wiltshire, Derbyshire, Kent, Leicestershire, Lincolnshire, West Yorkshire, Essex Police, as well as the College of Policing.

My range of pieces of research on police intelligence regulation; data protection and information sharing in public protection contexts (policing and health); and algorithmic justice issues in law enforcement contexts is set out in Appendix 2. Copies are of these pieces of research are either freely available online or are available on request. I would like to thank the Committee for any of their time spent considering my written evidence.

---

[4] See Yeung, K. A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework, 9th November 2018, from: https://rm.coe.int/draft-study-of-the-implications-of-advanced-digital-technologies-inclu/16808ef255 (accessed 08.04.2019)

[5] See Police Foundation, *Data-Driven Policing and Public Value*, 20th March 2019, from: http://www.police-foundation.org.uk/2017/wp-content/uploads/2010/10/data_driven_policing_final.pdf (accessed at 08.04.2019)

[6] See Liberty, *Policing by Machine*, 6th February 2019, from: https://www.libertyhumanrights.org.uk/policy/report-policing-machine (accessed at 08.04.2019)

Appendix 1

Advisory	Is the assessment made by the algorithm used in an advisory capacity? Does a human officer retain decision-making discretion? What other decision-making by human officers will add objectivity to the decisions (partly) based on the algorithm? *Care should be taken to ensure that an algorithm is not inappropriately fettering an officer's discretion, as natural justice and procedural fairness claims may well arise. Consider if supposedly advisory algorithmic assessments might in practice have an undue influence. If it is proposed that an algorithmic decision is automated and determinative then data protection rights in regard to automated decisions will apply.*
Lawful	On a case-by-case basis, what is the policing purpose justifying the use of algorithm: both its means and ends? Is the potential interference with the privacy of individuals necessary and proportionate for legitimate policing purposes? In what way will the tool improve the current system and is this demonstrable? Are the data processed by the algorithm lawfully obtained, processed and retained, according to a genuine necessity with a rational connection to a policing aim? Is the operation of the tool compliant with national guidance [that applies]? *The algorithm's proposed functions, application, individual effect and use of datasets (police-held data and third party data) should be considered against necessity, proportionality and data minimisation principles, in order to inform a decision to implement the tool. In relation to tools that may inform criminal justice disposals, regard should be given to the duty to give reasons. Logging these considerations can form the start of the requisite data protection/human rights (and public sector equality duty) impact assessment(s); <u>but these should be completed too.</u>*
Granularity	Does the algorithm make suggestions at a sufficient level of detail/granularity, given the purpose of the algorithm and the nature of the data processed? Is data categorised to avoid 'broad-brush' grouping and results, and therefore issues of potential bias? Do the potential benefits outweigh any data quality uncertainties or gaps? Is the provenance and quality of the data sufficiently sound? Consider how often the data should be refreshed. If the tool takes a precautionary approach in setting 'trade-offs', consider the justifications for this. *Consideration should be given to common problems in data analysis, such as those relating to the meaning of data, compatibility of data from disparate sources, missing data and inferencing. Consider how much averaging or blurring has already been applied to inputs (e.g. postcode area averages).*
Ownership	Who owns the algorithm and the data analysed? Does the force need rights to access, use and amend the source code and data analysed? How will the tool be maintained and updated? Are there any contractual or other restrictions which might limit accountability or evaluation? How is the operation of the algorithm kept secure? *Consider intellectual property ownership, maintenance of the tool and whether open source algorithms should be the default. When drafting procurement contracts with third party software suppliers (commercial or academic), require disclosure of the algorithmic workings in a way that would facilitate investigation by a third party in an adversarial context if necessary. Ensure the force has appropriate rights to use, amend and disclose the tool and any third party data. Require the supplier to provide an 'expert' witness/evidence of the tool's operation if required by the force.*
Challenge	What are the post-implementation oversight and audit mechanisms e.g. to identify any bias? Where an algorithmic tool informs criminal justice disposals, how are individuals notified of its use (as appropriate in the context of the tool's operation and purpose)? *The results of any algorithmic analysis should be applied in the context of appropriate professional codes and regulations. Consider whether the application of the algorithm requires information to be given to the individual and/or legal advisor. Regular validation and recalibration of the system should be based on publicly observable (unless non-disclosable for policing/national security reasons) scoring rules.*
Accuracy	Does the specification of the algorithmic tool match the policing aim and decision policy? Can the stated accuracy of the algorithm be validated reasonably periodically? Can the percentage of false positives/negatives be justified? How was this method chosen as opposed to other available methods? What are the consequences of inaccurate forecasts? Does this represent an acceptable risk (in terms of both likelihood and impact)? Is the algorithmic tool deployed by those with appropriate expertise? How are results checked for accuracy, and how is historic accuracy fed back into the algorithm for the future? Can forces understand how inaccurate or out-of-date input data affects the result?
Responsible	Would the operation of the algorithm be considered fair? Is the use of the algorithm transparent (taking account of the context of its use), accountable and placed under review alongside other IT developments in policing? Would it be considered to be used in the public interest and to be ethical? *It is recommended that ethical considerations, such as consideration of the public good and moral principles (so spanning wider concerns than legal compliance) are factored into the deployment decision-making process. Administrative arrangements such as an ethical review committee incorporating independent members could be established for such a purpose. PCCs should be briefed on the adoption/oversight of tools.*
Explainable	Is appropriate and intelligible information available about the decision-making rule(s) and the impact that each factor has on the final score or outcome? Is the force able to access and deploy a data science expert to explain and justify the algorithmic tool? The latest methods of interpretable and accountable machine learning systems should be considered and incorporated into the specification as appropriate. This is particularly important if considering deployment of 'black box' algorithms, where inputs and outputs are viewable but their internal workings are opaque.

Appendix 2

1. J. Grace, 'Human rights, regulation and the right to restrictions on algorithmic police intelligence analysis tools in the UK', available online as a draft paper at: http://ssrn.com/abstract=3303313

2. J. Grace, 'Algorithmic impropriety in UK policing?', (2019) *Journal of Information Rights, Policy and Practice*; Vol, 3 Issue 1: https://jirpp.winchesteruniversitypress.org/articles/abstract/23/

3. J. Grace, 'The policy shift towards more intensive monitoring of domestic violence offenders', March 2019, (2019) *Journal of Information Rights, Policy and Practice;* Vol. 3 Issue 1: https://jirpp.winchesteruniversitypress.org/articles/abstract/26/

4. M. Duggan & J. Grace, 'Assessing vulnerabilities in the Domestic Violence Disclosure Scheme', (2018) 2 *Child and Family Law Quarterly* 145

5. M. Oswald, J. Grace, S. Urwin & G. C. Barnes (2018) Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental' proportionality, *Information & Communications Technology Law*, Volume 27, 2018 - Issue 2, 223-250.

6. J. Grace, 'Countering extremism and recording dissent: Intelligence analysis and the Prevent agenda in UK Higher Education' (2017), *Journal of Information Rights, Policy and Practice* Vol. 2(2) (online)

7. J. Grace, 'Lessons on legislating for public protection information sharing: A case commentary on *Christian Institute v Lord Advocate* [2016] UKSC 51', *Journal of Information Rights, Policy and Practice*, (2017) Vol. 2, Issue 1, online

8. M. Oswald and J. Grace, 'Intelligence, policing and the use of algorithmic analysis: a freedom of information-based study, *Journal of Information Rights, Policy and Practice* (2016) 1(1) (online)

9. J. Grace and Oswald, M. '"Being on our radar does not necessarily mean being under our microscope": The Regulation and Retention of Police Intelligence', (2016) 22(1) *European Journal of Current Legal Issues* (online)

10. J. Grace 'The nature of spent convictions and the common law basis of the Domestic Violence Disclosure Scheme: Limiting the effectiveness of Clare's Law?', (2015) 21(2) *European Journal of Current Legal Issues* (online)

11. J. Grace, (2015) 'Better information sharing, or 'share or be damned'?', *The Journal of Adult Protection*, Vol. 17 Iss. 5, Online

12. J. Grace, 'Clare's Law, or the national Domestic Violence Disclosure Scheme: The contested legalities of criminality information sharing', *Journal of Criminal Law* (2015) 79(1) 36-45

13. J. Grace, 'The surveillance of 'risky subjects': adiaphorisation through criminal records, and contested narratives of stigma', (2014) 2(2) *Birkbeck Law Review*, 279-292

14. J. Grace, 'Old convictions never die, they just fade away: The permanency of convictions and cautions for criminal offences in the UK', *Journal of Criminal Law* 2014, 78(2), 122-136

15. J. Grace, 'The Information Governance Review and the new legal framework for informatics', *British Journal of Healthcare Management* (2014) 20(1), 40 - 44

16. J. Grace, 'Privacy, stigma and public protection: A socio-legal analysis of criminality information practices in the UK', *International Journal of Law, Crime and Justice* 41 (2013) 303-321

17. J. Grace, 'A broad discretion to share patient information for public protection purposes: Statutory powers of the NHS Commissioning Board', *Journal of Medical Law and Ethics*; (2013) 1(1), 77-83.

18. J. Grace & M. Taylor, 'Disclosure of confidential patient information and the duty to consult: The role of the Health and Social Care Information Centre', *Medical Law Review* (2013) 21(3), 415-447

19. J. Grace, '"Too well-travelled", not well-formed? The reform of 'criminality information sharing' in England and Wales', *Police Journal* (2013) 86(1) 29-52.

SUBMISSION 3

Written evidence submitted by SAS (DIG0015)

1. Introduction

1.1 The global leader in innovative analytics software and services, SAS transforms data into insight and knowledge that enables governments to identify what's working and fix what isn't. Whether boosting departmental performance or accelerating the deficit reduction, SAS analytics help the UK Government to improve service delivery to citizens, reduce costs and make better decisions faster.

1.2 SAS has over four decades of experience working with the public sector and is trusted and used by thousands of civil servants and hundreds of public sector customers across the UK. The company is the world's largest privately held software business and its analytics software is used by most of the Fortune 500 and by customers at more than 83,000 sites globally. Combined, this experience gives SAS unique insight as a partner to Government.

2. SAS working with government

2.1 SAS Institute works as a software partner and domain expert across almost all government departments. This includes:

- HMRC (tackling non-compliance, fraud and debt and providing the technology behind the CONNECT system);
- DWP (combatting benefit fraud and providing engagement and communication on appropriate entitlements);
- Ministry of Justice (situational intelligence in Prisons);
- NHS Digital (Health Informatics);
- The Bank of England (stress testing the City);
- Financial Conduct Authority (identifying money laundering and insider trading);
- Home Office (Detention Centres);
- Met Police (Criminal Intelligence);
- HM Treasury (Budget decision making);
- Army (HQ's MIB programme including paperless eCAB);
- DVSA (commercial vehicle stop and inspect regime) and
- wider work of ONS and National Statistics across England, Wales and Scotland.

2.2 SAS applaud the leadership demonstrated by the Civil Service in bringing together multiple government bodies and departments to ensure that holistic discussion is undertaken to identify savings & efficiencies.

2.3 SAS is currently seeking to develop and deploy its capability to help government departments identify procurement fraud. Results with existing clients and programmes have been rapid and have identified at least 1% fraudulent spend, saving billions of pounds.

3. Digital Leadership

3.1 Investment now in analytics will **help central government break out of a reactive cycle** where poor performance necessities urgent cash bail outs for failing services, and can put in place a long-term strategy for improved public services.

3.2 Government departments supported by multiple different agencies means there is **significant opportunity to use technology to improve information flow across government.** This will produce a better understanding and management of risk that result in faster and improved decision making.

3.3 Embracing digital technologies like analytics and AI is key to unlocking the digital dividend and should now be considered an essential element of civil service capability. This would help government cope with upcoming challenges without sacrificing improved service delivery.

3.4 The Industrial Strategy recognised as one of its grand challenges the opportunities presented by AI and set an objective of putting the UK at the forefront of the AI and data revolution. The potential contribution of AI to the economy is also highlighted, potentially adding £232bn by 2030. SAS is determined to work with government to maximise these opportunities and deliver the four priorities identified in the strategy. We are especially keen to engage with the Government Office for AI and play a leading role in an industry-led AI Council.

4. Responding to specific points raised in Terms of Reference

4.1 The progress of Government Digital Services, the areas where further development is particularly needed, and how well the UK compares with other countries.

4.1.1 SAS welcome the UK Government commitment to delivering digital services and leadership to date. However, competing demands and limited resources means that government is currently facing a bottleneck to make further advances.

4.1.2 SAS can share examples of how other large private and public sector oragnisations around the world are innovating and transforming service delivery through AI to help achieve this goal. Unlocking the potential of data-enabled technologies can yield huge social and economic benefits for the UK. It has been estimated that Artificial Intelligence (AI) alone could add an additional USD $814 billion (£630bn) to the UK economy by 2035, increasing the annual growth rate of GVA from 2.5 to 3.9%.[1] As such, government like other constituents in the UK has a massive amount to gain from using data-enabled technologies such as analytics and AI.

4.1.3 The key to unlocking the potential of data-enabled technologies for government as part of its digital transformation is trust. However, 2017 saw a record fall in trust across the institutions that shape our society, including government.[2] The findings from the Edelman Trust Barometer 2017 showed that trust in the UK Government stands at just 31 per cent.[3]

4.1.4 According to the OECD, this wider downturn in trust is in part due to nervousness from citizens about the way governments are using their personal information.[4] These findings are supported by other polls like KPMG's data trust deficit poll,[5] which showed a majority have a lack of trust in the public sector collecting and using data (37 per cent believing information security to be the biggest challenge).[6]

4.1.5 Conversely though, provision of better services by government, which data-enabled technologies deliver, actually builds trust in government. The OECD state that improving public services, in terms of access, quality and responsiveness, plays an important role in strengthening trust in government. They state that "service performance, citizen satisfaction and public trust are closely connected".[7]

4.1.6 So, on the one hand citizens have a lack of trust in government's data collecting and use. However, on the other, citizens want improved services that data-enabled technologies deliver; with this boosting trust in government as a result. This encapsulates the opportunity created in digital services.

4.1.7 Creating a 'spine' of citizen data underpinned by common data dictionaries for joined up, cross-departmental approaches to citizen service provision is already possible through tools like analytics. Such moves would also not require data to be stored in a central database either, as tools like SAS analytics can use 'snapshot' approaches to take the required data from existing internal and external silos of data. This cuts costs for governments from utilising existing tools. It also helps to calm citizen concerns around potential over harvesting of their data that such a centralised source of data could cause.

4.1.8 Whilst the civil service isn't short of digital leaders, the UK tends to be held back by the vast complexity of the machinery of government that has evolved. Established ways of working reduces the freedom and flexibility required for new services (especially cross-cutting ones) to be delivered rapidly.

4.1.9 A lack of faith in the data is present amongst civil servants for many good reasons but SAS believe that this can be addressed through collaborative efforts focused on improving the management and quality of the available data.

4.2 How well Government digital services are protected from cyber-attacks.

4.2.1 This is not appropriate for SAS to comment upon. Any intelligence we may have is shared directly with relevant departments and civil servants.

4.3 How well the Government Digital Service (GDS) has helped spread the use of digital services across government, including promoting the use of new technologies and uses of data.

4.3.1 At the recent Sprint conference, GDS showcased a variety of new capabilities that have been developed over recent months. They are a tremendous credit to the pioneering civil servants who have deployed them and provides a clear demonstration of their leadership for promoting the use of new technologies to support their endeavours.

4.3.2 In particular, the work in joining up the machinery of government from multiple departments to provide a citizen centric web application was outstanding. The application hides much of the legacy inter-departmental bureaucracy and should be applauded for the sheer force of will it must have taken to overcome the departmental obstacles (both real and perceived) to delivery.

4.3.3 It is vital that the learnings from such programmes are fed back into the policy functions of the departments concerned to ensure that lasting change endures within government processes. Perhaps there is a greater role for GDS to have responsibility for incorporating government (as a platform) into our non-digital delivery processes if we are to maximise efficiency and reduce the burden of administration.

4.3.4 We welcome the pragmatic stance adopted on the open source software policy and believe that a hybrid approach will ensure an appropriate level of flexibility. This pragmatism has also avoided the expenditure of civil service resource on well-intentioned but costly, time consuming and high-risk endeavours.

4.3.5 The learnings from this experiment do have merit and we encourage government to look for ways in which proprietary software tools, techniques, algorithms, data models and APIs can be

used across departmental boundaries to accelerate transformation. Such a collaborative approach will de-risk projects whilst reducing costs and SAS would welcome opportunities to foster closer working ties with users across governments around the world to assist GDS in this endeavour.

4.3.6 GDS has helped to create an opportunity for government to accelerate the pioneering adoption of AI for the benefit of its service to citizens, to deliver export opportunities and showcase examples of such capabilities and to enrich the lives of peoples across the globe through the ethical oversight that such leadership imparts.

4.4 The digital skills capacity in government departments and agencies, to be able to deliver effective digital services to the public and businesses.

4.4.1 SAS, coming from the world of education has long been a proponent of developing the relevant STEM skills and is a strong corporate sponsor of aligned courses. SAS promotes learning in statistics and data science from infant schools to master's programmes globally. The skills challenge will increase and a strategic solution is required.

4.4.2 SAS is keenly aware of the challenges and works closely with central government to offset these. We have developed departmental academies for our users to enable the civil service to retrain numerate staff with vital subject matter expertise to counter recruitment difficulties. We work with heads of profession to promote a progressive career path for data scientists to reduce the skills drain to industry caused low morale and pay disparity.

4.4.3 As local government starts to take on an increased role in delivery data and analytics will become of ever increasing importance to a sector that is ill prepared and equipped. SAS is ready to work with Government in the challenge to educate and train this vast cohort as the risks inherent in this domain are significant.

4.5 How well the UK Government and its agencies deploy their datasets to maximise their value for money, effectiveness and delivery of digital services.

4.5.1 SAS sees some great examples of data exploitation across government and also a significant number missed opportunities.

4.5.2 There are any number of causes of sub-optimal exploitation of data across government. There remains a risk-averse caution over data sharing despite GDS and Efficiency and Reform Group (ERG) efforts to break this down. This sometimes seems exacerbated by outsourcing contracts and complex commercial arrangements.

4.5.3 Radical use of data to reduce the size of the civil service remains counter intuitive to the career civil servant. Opportunities to automate manual processes using large data analysis fall foul of a desire for zero-defect outcomes – even when the status quo is often imperfect in its execution.

4.5.4 There is significant opportunity to use AI and analytics to improve information flow across the government departments. For example, in the justice system, better understanding and managing of risk will aid and protect staff, and provide critical insight in decisions affecting individual's pathway and treatment through the criminal justice system. By utilising the data held utilising the data held about prisoners and applying analytical method and modelling capabilities it is possible to reduce recidivism. This will also ensure that senior civil servants can focus on delivering key policy priorities.

4.5.5 There are multiple opportunities to monetise valuable data held by government. This will require strict guidelines and clear control mechanisms. The 2021 census provides the prospect of building an electronic population record that can enable policy simulation and modelling across government.

4.6 The extent to which Government datasets are made available to private-sector and academic service developers, and how well its 'open data' arrangements are operating.

4.6.1 This area is one that SAS has supported through certain initiatives within individual departments for many years. We still provide these services and the open data initiatives have not replaced the need for them. This suggests that they do not meet all requirements but we cannot comment further at the present time.

4.7 The implications and opportunities for GDS arising from Brexit, including areas where the nature of digital services may have to change.

4.7.1 The scale of the anticipated change presents a unique challenge and opportunity. The ability of digital technologies to reduce cost and build time for Government services will be tested, from regulatory changes to capacity issues, as roles are taken back from the EU into government departments post Brexit.

4.7.2 The period outlined for the UK Government's Transformation Strategy up until 2020 will need to deliver large and fast paced change.

4.7.3 There is a clear need to accelerate the uptake of advanced analytics solutions that can assist government in making more informed decisions more rapidly. Specifically, the impact of Brexit on immigration controls and trade flows must be addressed as soon as possible.

4.7.4 Artificial intelligence can not only help government in coping with Brexit pressures; but in both the immediate and near-term has the potential make Brexit a positive catalyst towards the UK Government achieving its goals of digital transformation to be more responsive to the needs of citizens through adopting more agile practices. In this briefing three key areas are explored:

- Analytical AI Innovation $(AI)^2$ offers new solutions to Brexit challenges, such as the NI Border.
- $(AI)^2$ will increase the capacity of Government to deliver a successful Brexit whilst increasing service delivery capacity.
- Increasing uptake of $(AI)^2$ to facilitate Brexit, will speed up delivery of the Government's goals of more agile practices and long-term digital transformation.

4.7.5 SAS is working with many countries globally, creating innovative ways to solve border issues similar to those seen with the Northern Ireland land border. SAS uses patented technology to assess if the unusual is actually suspicious and offers a way to significantly reduce the associated infrastructure at the border both for immigration and customs management. Its repertoire of AI capabilities enables machine learning to be employed taking full advantage of numerous SAS patented anomaly detection and monitoring engines within their border surveillance portfolio to help categorise threats and prioritise interventions.

4.7.6 Flexible solutions to Brexit capacity issues are required. They must tackle different scales of challenge without compromising other departmental service delivery. SAS can facilitate conversations with customers in Europe who have effectively joined up the data between their business enterprise and their workforce. In effect, this has enabled EU counterparts to BEIS and

DWP to offer a joined-up service to citizens and businesses to optimise their workforce and productivity.

4.7.7 It is also important to note that AI solutions can also turn the challenges that Brexit bring to capacity and service delivery into a positive, as Government realises the alignment between AI solutions and the critical long-term goal of digital transformation and creating a more agile process system. An agile approach to project management aims to encourage teams to build quickly, test what they've built and iterate their work based on regular feedback.

4.8 The implications for GDS following the move of its data policy and governance functions from the Cabinet Office to the Department for Digital, Culture, Media and Sport.

4.8.1 SAS have been keen to understand how best to support the UK Government across the two main domain areas in which we are global leaders – Artificial Intelligence and Data Analytics. We therefore welcome this move as our discussions within and around central government (taking account of opinions from Whitehall and the view from Westminster), seemed to be revealing a lack of strong political leadership in these areas.

4.8.2 GDS, being one of many Cabinet Office functions and so lacking a dedicated minister, has been confined latterly to leading important but mundane capability building initiatives. As the UK now seek to capitalise on these capabilities it is a timely and necessary move.

4.8.3 It is hoped that the DCMS broader portfolio, including the critical strands of data privacy and the establishment of a framework for ethical AI taken alongside their role in partnership with BEIS to support the wider information technology industry, will provide a focus for this area as it becomes ever more central to the UK economy, foreign policy and indeed to the everyday lives of this nation and the world in the coming decades.

5. Conclusion

5.1 The UK Government must continue to improve the way in which it makes spending decisions to improve public service delivery. Investment now in analytics and AI will help government achieve its goals of digital transformation and delivering more agile and efficient public services.

5.2 Analytics can help to deliver more comprehensive analysis of intra-governmental risks, robust and real-time evidence to inform decision making, the identification of current and future skills gaps as well as improved policy performance management systems that encourage flexibility and empower decision makers with detailed and holistic information.

5.3 The digital dividend that will be delivered by key innovations in analytics can ensure that government is able to handle future challenges whilst maintaining efficient service delivery.

Submitted September 2018

---

[1] 'Growing the artificial intelligence industry in the UK', October 2017, DCMS: https://www.gov.uk/government/publications/growing-the-artificial-intelligence-industry-in-the-uk/executive-summary

[2] 'Edelman Trust Barometer 2017 UK Findings', January 2017, Edelman: https://www.edelman.co.uk/magazine/posts/edelman-trust-barometer-2017-uk-findings/

[3] ibid.,

[4] 'How Better Governance Can Help Rebuild Public Trust', June 2017, OECD: http://dx.doi.org/10.1787/9789264268920-en

[5] 'The data-trust deficit;, May 2017, KPMG: https://home.kpmg.com/xx/en/home/insights/2017/04/the-data-trust-deficit.html

[6] Ibid.

[7] How Better Governance Can Help Rebuild Public Trust', June 2017, OECD: http://dx.doi.org/10.1787/9789264268920-en

SUBMISSION 4

The Committee on Standards in Public Life

Artificial Intelligence (AI) and public standards review

Written submission by:

Marion Oswald
Solicitor (non-practising)
Senior Fellow in Law and Director of the Centre for Information Rights, University of Winchester
Associate Fellow, Royal United Services Institute
[redacted]

Introduction

1. This submission is based upon my research into algorithm-assisted decision-making in the public sector, and in particular within policing, and also upon my experience as Chair of the West Midlands Police and Crime Commissioner and West Midlands Police data ethics committee.[1]

2. The submission comments upon the Committee's terms of reference and makes a number of suggestions regarding the application of the Seven Principles of Public Life to algorithm/AI-assisted public sector decision-making, including:

- *Selflessness*: AI technology is often experimental. Challenges arise in determining whether deployment within front line public services is in the public interest and if the human rights necessity and proportionality tests are made out, suggesting the need for clear policy frameworks and oversight structures.
- *Objectivity*: public servants require information in order to uphold this standard, both as regards the commissioning and procurement of AI, and in relation to AI-assisted decision-making. Additional rights for the public sector party (as set out below at 7. and 9.) should be included as standard and compulsory conditions in any tender. Scientific validity and relevance standards are needed to enable public servants to judge whether AI outputs are the 'best evidence'. Furthermore, the principles of administrative law can be reframed to help preserve objectivity in an algorithm-assisted environment.
- *Accountability*: the decision-making process, of which the AI tool is part, must preserve the human discretion to assess 'un-thought of' relevant factors – such as positive factors that may rely on uncodified knowledge - and to assess whether the question or decision is the one for which the AI was designed. AI and machine learning tools should not be inserted into a process that requires the exercise of discretion by a public authority where the tool prevents that discretion.

Are existing frameworks and regulations sufficient?

3. There is evidence of increasing interest in the use of data analytics and machine learning tools to improve UK public sector decision-making and service provision. This follows the trend already seen in the United States. For instance, my research has analysed the deployment of the HART decision-support tool within Durham Constabulary, the first deployment of a machine learning decision-support tool in an operational capacity.[2] West Midlands Police has created its own internal Data Analytics Lab, with the ethics committee created in parallel to advise the Chief Constable and PCC on its projects. Cardiff University's Data Justice Lab recently reported on the use of data scoring in public services, providing a useful overview of some

[1] https://www.westmidlands-pcc.gov.uk/transparency/ethics-committee.
[2] Marion Oswald, Jamie Grace, Sheena Urwin & Geoffrey C. Barnes (2018) 'Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental' proportionality' *Information & Communications Technology Law*, DOI: 10.1080/13600834.2018.1458455.

deployments of algorithmic methods across the UK public sector.[3] Oxford University's report into Data Science for Local Government concluded that use of data science within this sector was at a 'nascent' stage.[4]

4. It has hard to say with any certainty how standards will be affected in particular contexts (whether positively or negatively) by the widespread introduction of AI and machine learning into public sector decision-making. Much commentary has highlighted potential risks and issues, particularly relating to the impact on individual rights. There tends to be a divide between those focused upon strengths and opportunities of data science and those who stress the risks and issues. Within the policing context, my jointly authored research has drawn attention to the limited evidence base on the efficacy and efficiency of different systems, their cost-effectiveness, their impact on individual rights and the extent to which they serve valid policing aims.[5] Meijer & Wessels argue for more research into how predictive models work in practice (to see if drawbacks actually occur),[6] and the recent US report on predictive policing by the Partnership on AI found that more research is required on how data-enabled risk assessment tools inform human decisions, in order to determine what forms of training will support principled and informed application of these tools, and where gaps exist in current practice.[7]

5. A number of legal frameworks are applicable to the development and deployment of AI within the UK public sector, including:

- Data protection (including the right not to be subjected to automated decision-making, and the principles regarding sensitive personal data, accuracy, retention, security and data protection impact statements);
- The Public Sector Equality Duty (the 'due regard' standard under the s149 Equality Act 2010, and other EA 2010 standards);
- Obligations pursuant to the ECHR, taking effect through s6 Human Rights Act 1998;
- Administrative law principles applicable to lawful public sector decision-making.[8]

6. These legal frameworks are primarily principles-based, meaning that often difficult judgements are required on a case-by-case basis regarding such issues as the justification and relevance of data inputs, and the necessity and proportionality both of the data analysis and the way in which the output is then used. Furthermore, the lack of guidance frameworks regarding methods of 'testing' these technologies, particularly within operational environments,[9] and the absence of clear scientific standards by which to judge the validity of the outputs,[10] add considerably to the challenge.

---

[3] Dencik, L., Hintz, A., Redden, J. and Warne, H. (2018) Data Scores as Governance: Investigating uses of citizen scoring in public services. Research Report. Cardiff University. See also Fieke Jansen, Working Paper: Data Driven Policing in the Context of Europe, 7 May 2019 https://datajusticelab.org/2019/05/08/new-report-from-datajustice-project-on-data-driven-policing/.

[4] https://smartcities.oii.ox.ac.uk/data-science-for-local-government-report/.

[5] Alexander Babuta, Marion Oswald and Christine Rinik (2018) 'Algorithms, Predictive Policing and Criminal Justice Decision-Making' *Whitehall Report* (published by the Royal United Services Institute on 21 September 2018) https://rusi.org/publication/whitehall-reports/machine-learning-algorithms-and-police-decision-making-legal-ethical.

[6] https://www.tandfonline.com/doi/full/10.1080/01900692.2019.1575664.

[7] https://www.partnershiponai.org/report-on-machine-learning-in-risk-assessment-tools-in-the-u-s-criminal-justice-system/.

[8] Marion Oswald (2018) 'Algorithmic-assisted decision-making in the public sector: framing the issues using administrative law rules governing discretionary power' in 'The growing ubiquity of algorithms in society: implications, impacts and innovations' issue of Philosophical Transactions of the Royal Society A.

[9] n5.

[10] Marion Oswald 'Technologies in the twilight zone: Early lie detectors, machine learning and reformist legal realism' (April 7, 2019). Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3369586.

7. The involvement of private sector commercial organisations in the development and provision of AI and data analysis tools for use in public sector environments (for instance the involvement of Accenture in the proof-of-concept stage of the police National Data Analytics Solution project, and IBM in the National Law Enforcement Data Service) raises additional considerations over and above those raised in a standard software procurement scenario. These include the need for the public sector body to have:

- knowledge of, confidence in, and appropriate access to, the *training datasets* with which the model was built;
- appropriate *access to the algorithmic workings* in a way that would facilitate investigation and questioning by a third party in an *adversarial* context;[11]
- rights to *use, amend and disclose* the tool, its workings and the input datasets where required for the purposes of legitimate public sector decision-making;
- regular rights of *audit, testing and validation*, including but not limited to, in relation to potential for bias, and false positive and false negative rates;
- rights to require the *updating* of the model, including by way of removing or adding input factors;
- access to an *'expert' witness* from the commercial supplier to provide evidence as to the tool's operation.[12]

Building in standards

Commissioning AI

8. Assertions of commercial confidentiality should not be accepted as an insurmountable barrier to appropriate rights of access to the tool and its workings for the public sector body, particularly where the tool's implementation will impact fundamental rights. Government procurement contracts relating to AI and machine learning should not only include source code escrow provisions, but rights for the public sector party as set out in 7. above *as standard* and as *compulsory conditions for any tender*. Furthermore, the above requirements apply equally to the commissioning of a higher education or third sector body to develop a machine learning tool.

9. In the development process, it can often be the case that a third party provider is permitted a large degree of autonomy regarding selection of the input datasets, method of data analysis and communication of the output via the human/computer interface. Terzis argues that 'as computational technology becomes more complex, the client will be gradually withdrawing from specifying the developer's conduct and the developer, in turn, will be acquiring a degree of contractual freedom qualitatively similar to the lawyer's.'[13] In addition to the rights suggested in 7. above, this suggests the need for frameworks to:

- ensure that the *client retains decision-making authority over the purpose* of the tool (the question that the analysis is to answer) and the way that the output is presented to ensure that the tool is not overreaching. (I have argued previously that 'we cannot always assume that the forecast or classification represents the only or main factor on which the 'rightness' or 'wrongness' of the overall decision is to be judged. Doing so may risk changing the question that the public sector decision-maker has to answer.'[14]);

[11] See the 'Algocare' model (Oswald and Grace, 2017).
[12] n11.
[13] Petros Terzis 'The Reasonable Coder' https://robots.law.miami.edu/2019/wp-content/uploads/2019/03/The-Reasonable-Coder.pdf Presented at *WeRobot 2019*, April 2019.
[14] n8.

- give rights to the client over the *input datasets*: data may be 'relevant' from a statistical perspective (in that the data may increase statistical accuracy) but may be irrelevant or disproportionate from a legal perspective to the public sector question to be answered;[15]

- enhance the standard requirement that services should be provided with 'reasonable care and skill' to reflect the overall *legal and operational environment* in which the tool will be deployed. (For instance, in the criminal justice context, providers could be expected to take account of the proportionality of using certain datasets relating to early years offending);

- require providers to produce *algorithmic risk assessments* addressing matters such as accuracy, bias and security;[16]

- require providers to make changes to the model if requested by an *oversight or review body*, including an ethics committee;

- provide *authoritative expert guidance* to public sector bodies on the *validity and limitations* of particular models and methods. (Machine learning has been described as 'alchemy,'[17] common statistics favoured to demonstrate a model's predictive validity have been argued to be misleading and uninformative in the context of offender risk management due to the high margins of error often involved,[18] and focus solely on predictive data analysis without incorporating causal expert knowledge is said to be unlikely to lead to better decisions.)[19] I have argued for the development and application of appropriate '*scientific validity*' and *relevance* standards for AI and machine learning, constructed for particular criminal justice contexts (investigative, offender management, risk-assessment and so on), and to include the presentation of results (to ensure these are not presented as 'something more'), compliance with which would be a red-line.[20] This guidance could also assist public sector bodies in crafting appropriate contract specifications and standards for AI tools, and advise on how performance should be tested and judged.[21]

Deployment

10. The responsible development of algorithmic tools could provide new and potentially 'better' solutions for public sector problems (thereby contributing to best practice and opportunities for cross-sector openness and learning). This might particularly be the case in situations which currently involve clouded, non-augmented decision-making. Innovation with real data is therefore required in order to test those solutions, and innovation implies a degree of uncertainty about outcome and whether deployment will achieve a legitimate aim.

---

[15] n8.

[16] See the proposed Algorithmic Accountability Act introduced in the US Senate which aims to tackle potential algorithmic discrimination and unfairness by way of impact assessments: https://www.theregister.co.uk/2019/04/10/us_ai_bill/.

[17] Matthew Hutson, 'AI researchers allege that machine learning is alchemy' May 3, 2018 *Science*.

[18] Cooke D.J. and Michie C. (2014) 'The generalizability of the Risk Matrix 2000: On model shrinkage and the misinterpretation of the area under the curve' *Journal of Threat Assessment and Management* 1(1) 43.

[19] Miguel A. Hernán, John Hsu & Brian Healy (2019) A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks, CHANCE, 32:1, 42-49, DOI: 10.1080/09332480.2019.1579578.

[20] n10.

[21] Taking a recent paper as an example, 'Forecasting Knife Homicide Risk' by Massey and Sherman, https://link.springer.com/article/10.1007/s41887-019-00034-y although a correlation between locations of non-fatal stabbings and future knife homicide is claimed (on 2 years of data), the results include high false positives rates, potentially particularly problematic if increase in stop and search might result from the deployment of this model.

11. Uncertainty around outcome therefore raises questions around the necessity and proportionality of using group data to build an AI tool, for instance the re-purposing of criminal record and intelligence data about many individuals as training datasets for a machine learning predictive model. Although the individuals whose data have been analysed may never themselves interact with the tool, the 'echo' of their data informs the way the model operates and how it categories other individuals. If data on which training of the tool was based are subsequently deleted, amended or found to be inaccurate or flawed, this could have significant implications for the tool from both a legal and practical perspective.

12. Bearing these uncertainties in mind, in *Oswald, Grace, Urwin and Barnes* (2018),[22] we propose the implementation of what we have called an 'experimental' proportionality approach. Our approach would permit the use of unproven algorithms in the public sector in order that benefits and harms can be fully explored, but in a controlled and time-limited way, with the proportionality subject to a further review on a stipulated future date (so a similar aim to a 'sunset' clause in legislation).

13. The concept of 'experimental' proportionality would encapsulate two elements:

i) giving the 'benefit of the doubt' to the public sector body where it is not yet possible to determine with any certainty the balance or imbalance of benefits and disadvantages in relation to the new AI technology. This 'benefit of the doubt' would only apply if the public sector body can demonstrate a baseline connection to a legitimate aim and that the outcomes and benefits (even if these are as yet theoretical or only foreseen) are rationally connected to that aim and, based on the knowledge available, a reasonable belief that there is not an excessive cost to human rights; and

ii) a formal procedure (available to regulators, oversight bodies and the High Court in judicial review proceedings) so that the benefits and harm risks, and hence the proportionality of the particular use of the algorithm, can be ordered to be reviewed in another hearing after a period of time. The role of a suitable senior officer (e.g. the SIRO), aware of the algorithms detail, to interpret individual results and ensure that contextual and causal factors are considered cannot be underplayed in this proposed experimental stage.

14. We appreciate that the first stage of experimental proportionality is highly dependent upon there being confidence in the decision-making rigour of the public body: hence our linked proposal designed to contribute to such decision-making rigour, a guidance framework called 'Algo-care.'[23]

15. Furthermore, in *Babuta, Oswald and Rinik* (2018), we conclude that a new regulatory framework is needed within policing, one which establishes minimum standards around issues such as transparency and intelligibility, the potential effects of the incorporation of an algorithm into a decision-making process, and relevant ethical issues. A formalised system of scrutiny and oversight, including an inspection role for Her Majesty's Inspectorate of Constabulary and Fire and Rescue Services, is necessary to ensure adherence to this new framework.[24]

Ethics processes

16. The promotion of ethical principles and guidelines has been gaining traction,[25] although many of these initiatives can be criticised for a high level of abstraction, limited consideration of existing legal and regulatory regimes, and lacking any enforcement or oversight mechanisms. By contrast, I am involved in two

[22] n2.
[23] Copy provided with this submission.
[24] n5.
[25] Such as the EU ethics guidelines for trustworthy AI: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

public sector ethics committees (as a member of the National Statistician's Data Ethics Advisory Committee[26] (NSDEC) and Chair of the West Midlands Police and Crime Commissioner and West Midlands Police Ethics Committee). Both these committees operate in accordance with terms of reference, and review submissions against specified principles, which include legal compliance. Both committees have a commitment to transparency, with papers and minutes published online (subject to any necessary operational confidentiality).

17. The West Midlands committee's terms of reference tasks the committee with tracking a project from development to deployment, as it is anticipated that unforeseen consequences could occur when a project moves from the development stage to operational roll-out, with the PCC and Chief Constable required to respond to the Committee's feedback and provide reasons for any disagreement with the Committee's recommendations. Based on the proceedings of the NSDEC since 2015, the Statistics Authority has developed self-assessment and 'precedent' administrative processes, allowing researchers to assess projects in advance of full ethical review by NSDEC and compare new proposals against projects previously approved by NSDEC.[27]

18. Although entitled 'ethics' committees, the remit of these bodies is not in fact narrowly defined; they could be said rather to be oversight committees, testing proposals against the 'public good', and providing the benefit of a 'fresh pair of eyes.' The structure of these bodies might usefully be further studied in order to provide a template that could be used more widely within the public sector for oversight of the deployment of AI.

Implications for the Principles of Public Life

19. *Selflessness*: AI technology is often experimental. Challenges arise in determining whether deployment within front line public services is in the public interest,[28] and if the human rights necessity and proportionality tests are made out. Clear policy frameworks and oversight structures are required to support the testing, assessment and monitoring of AI in operational environments.

20. *Objectivity*: public servants require information in order to uphold this standard, both as regards the commissioning and procurement of AI, and in relation to AI-assisted decision-making. Rights for the public sector party as set out in 7. above should be included as standard and compulsory conditions in any tender. Scientific validity and relevance standards are needed to enable public servants to judge whether AI outputs are the 'best evidence'. Furthermore, the principles of administrative law concerned with the exercise of discretionary power (the duty to give reasons, relevant and irrelevant considerations, and the rules against fettering discretion and against bias) can be reframed to help preserve objectivity in an algorithm-assisted environment.[29]

21. *Accountability*: This Principle upholds the need for a human to remain ultimately accountable for the influence of an AI-assisted process. My reframing of the relevant administrative law principle supports this: The decision-making process, of which the algorithmic tool is part, must preserve the human discretion to assess 'un-thought of' relevant factors – such as positive factors that may rely on uncodified knowledge - and to assess whether the question or decision is the one for which the algorithm was designed. AI and machine learning tools should not be inserted into a process that requires the exercise of discretion by a public

[26] https://www.statisticsauthority.gov.uk/about-the-authority/committees/nsdec/.
[27] https://www.statisticsauthority.gov.uk/about-the-authority/committees/nsdec/data-ethics/.
[28] For instance, the *Impact Evaluation Summary of the Allegheny Family Screening Tool* of the predictive risk model implemented by Allegheny County in 2016 indicates mixed results: while identification of children in need for further intervention increased, there were no decreases in re-referral rates for children screened out, and no evidence of greater screening consistency https://www.alleghenycountyanalytics.us/wp-content/uploads/2019/05/Impact-Evaluation-Summary-from-16-ACDHS-26_PredictiveRisk_Package_050119_FINAL-5.pdf.
[29] n8.

authority where the tool prevents that discretion; either because all of the factors relevant to the decision cannot be included, or required elements of the decision itself cannot be appropriately codified into, or by, the algorithm.[30] Practical considerations, in particular design of the human-computer interface, the avoidance of unnuanced framing of results (such as 'traffic-lighting' of risk levels), and organisational culture and processes, will be crucial to the upholding of this Principle.

22. *Openness*: The incorporation of an AI tool into a decision-making process may come with the risk of creating 'substantial' or 'genuine' doubt as to why decisions were made and what conclusions were reached, both for the subject of the decision and for the decision-maker themselves.[31] Consideration should be given to the circumstances in which reasons for/an explanation of the output may be required. These may include: to determine whether the data on which the machine learning algorithm was trained match the circumstances of the current situation; the identification of situations where the output is likely to be flawed; where individual rights and freedoms are under consideration. The properties or granularity that should be provided by an explanation will be dependent upon the context, the particular user requiring the explanation and the likely weight of the outcome that the algorithmic output informs.[32] The results of these considerations should feed into the contractual requirements when commissioning a third party provider.

23. *Honesty*: in order to support this principle, organisational culture, policies and processes should support public servants in taking legitimate decisions contrary to the algorithmic recommendation.

24. *Leadership*: projects should not be driven by data science at all costs. Public servants should be prepared to walk away from experimental AI where there is no clear benefit to the public task and where the potential infringement of individual rights cannot be shown to be necessary and proportionate for legitimate public purposes. In some cases, this may require the 'writing off' of investment.

Marion Oswald

9 May 2019

The following publications are included with this submission for reference:

Marion Oswald, Jamie Grace, Sheena Urwin & Geoffrey C. Barnes (2018) 'Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental' proportionality' *Information & Communications Technology Law*;

The 'Algocare' guidance framework;

Marion Oswald (2018) 'Algorithmic-assisted decision-making in the public sector: framing the issues using administrative law rules governing discretionary power' in 'The growing ubiquity of algorithms in society: implications, impacts and innovations' issue of *Philosophical Transactions of the Royal Society A*;

Marion Oswald 'Technologies in the twilight zone: Early lie detectors, machine learning and reformist legal realism' (April 7, 2019).

[30] n8.
[31] n8.
[32] n8.

SUBMISSION 5

medConfidential note on AI and public standards

Summary

The UK prides itself on the principles of democracy and the rule of law. Those in public life are already required to show how they comply with the Nolan Principles of Public Life (for conduct), and the Bingham Principles on the Rule of Law (for decision making), the civil service code overlaps and covers some adjacent areas. That is no different because of the use of Algorithms, Data, or Artificial Intelligence,[1] or whatever comes after today's ADA.

New standards must be grounded in the long traditions of academic and practical work that underpins principled frameworks and practice. Standards applied to explaining decision making in automated systems should be derived from the Bingham Principles directly, and for policy processes from the Nolan Standards – especially around objectivity, fairness, and transparency – rather than be retrofitted off current practice. Human practice being an institutional norm which assumes that the status quo is not worth the effort to improve it, which has always been the argument of the status quo, and not that of innovators. It is a sign of the political seductiveness of these new technologies and their corporate advocates that the narrative has changed, embodying the view that the powerful should never be challenged with principles, but should be allowed to do what is easiest for them. The 'ethics' frameworks are a cargo cult of industry where profit and contracts are paramount, externalities are ignored, and caveat emptor.

A choice to make use of Algorithms, Data, or Artificial Intelligence changes neither the standards upon which decisions should be made, nor the consequences of failing to do so. Accountability and Openness are principles replicated in the Civil Service Code, and also replicated in the Data Protection Act (as accountability and fairness/transparency).

No matter how they are made, the Principles of the Rule of Law[2] as identified by Lord Bingham[3], alongside the Principles of Public Life as identified by Lord Nolan, must apply to and guide all decisions by public bodies.

There is no "AI exception" to the Principles of Public Life – nor to the Principles of the Rule of Law – in exactly the same way, and for the same reason, that there is no "typewriter" exception, nor "electricity" exception. Whatever the technology the Principles should still be applied. (There is no "AI exception" in the Data Protection Act either.)

The amount of 'harm' a single civil servant could previously effect in a single decision was largely limited to a single letter – the single swipe of a pen. With increasing use of algorithms and data, a single run of an algorithm on some data can result in adverse consequences for millions of citizens. The effects of a flawed decision are far, far greater, should the standards of public life be misapplied.

---

[1] A collection of elements the Ada Lovelace Institute helpfully shortens to "ADA" in its roadmap: https://www.adalovelaceinstitute.org/nuffield-foundation-publishes-roadmap-for-ai-ethics-research/
[2] https://binghamcentre.biicl.org/schools/ruleoflaw
[3] The Rule of Law, Tom Bingham. ISBN: 8601400310878

The Committee was set up to ensure that standards of public life are maintained, standing in defence against those shortcomings in human nature that tend to desire lower standards – whether for the sake of short-term 'innovation', sheer expediency, or the much rarer case of private profit.[4] One thing that does change is that the effectiveness of this defence becomes inherently measurable when the process is driven by AI and data.

As the use of algorithms, data, and AI moves forward, failures to uphold standards and principles will become measurable from the outcomes of decisions. This premise has already been accepted by Government, in the context of digital court outcomes.

Over time, all algorithmic, data, or AI decisions will be made understandable[5] – and clear economic and technological advantages arise from the UK's broadly known and understood standards for decision-making, compatible with both the Nolan Principles of Public Life and the Principles on the Rule of Law identified by Lord Bingham. The message from public bodies to every company wishing to provide services to them must be: here are the known standards, show us how you meet them.

Building trustworthiness in the use of algorithms, data, and AI

Technology developed to meet these standards can then be exported worldwide, re-usable by anyone who wishes to show that they comply with the Rule of Law and the highest public standards of decision-making.

Every system that uses data, algorithms, and AI should be expected to show how it satisfies its obligations – whether those be legal obligations under the Equality Act, obligations to the Civil Service code and standards on public life by decision makers, or transparent public statements to citizens.

Baroness O'Neill has written widely on the notion of trustworthiness[6] in institutions, underscoring the fact that the method of decision-making – whether by a human, a machine, or the throw of a dice – should not affect the principles of transparency. The method of decision-making – and indeed any outsourcing, whether that be to companies or machines – will not affect the public's perception of those in public life who will be held responsible for the real world outcomes.

In the First Report of the Committee, a strong case was made for Standards to cover Special Advisors in the same way they would cover Ministers and relevant civil servants. The same arguments now apply to AI, for the same reason that the current Terms of Reference for the Committee are related to "paid for by public funds".

---

[4] The comments in the Committee's First Report about this being a rare exception (then) also apply here, but public perception may diverge from what remains a generally unknown practice.
[5] https://medconfidential.org/2018/ai-and-demonstrations-of-political-power/
[6] Onora O'Neill (2018) Linking Trust to Trustworthiness, International Journal of Philosophical Studies, 26:2,293-300, DOI: 10.1080/09672559.2018.1454637
https://www.tandfonline.com/doi/abs/10.1080/09672559.2018.1454637

Existing methods to demonstrate how standards are followed

Tribunals (and other courts) could be described as the 'exception handling' system of the rule of law and public life[7] – for when something has gone wrong, and parties disagree, or when someone has seriously breached the law.

In order to understand the impact of its digital tribunals programme, and encouraged by Parliament, HMG has committed to collect data on outcomes and protected characteristics.[8] HMCTS Expert Advisor[9] Dr Byrom's paper[10] catalogues the 13 attributes necessary to monitor access to justice, as required under existing legislation – defining each of them clearly and distinctly in terms of law, standards, and principles.

As HMCTS implements, iterates, and evolves the approach identified by Dr Byrom, there will be evidence of practical differentials in access to justice and the rule of law, measured first at the entry and exit points of the courts and tribunal system.

Recommendation: All uses of AI should demonstrate their compliance with the openness, honesty, and fairness Principles of Public Life, as part of their Data Protection Impact Assessment, which should include assessing Equality Act protected characteristics at the entry and exit points for any use of algorithm or AI. This should also be applied to the justifications required under s35 of the Digital Economy Act.

The Nolan Principles in the modern world

It has always been the case that human beings fall short of principles, eventually. Much of the justice system (and civil society) is about minimising the damage when people justify their actions in retrospect – no tyrant ever failed to justify their crimes. People only get caught in the most egregious cases where breaches are institutionalised, e.g. redlining – and institutional discrimination is very difficult to actually prove (...sexism…) at a human and individual level. It has to be shown systematically.

In an automated world, everything is systematic, so it is far far easier. Every algorithm used by public bodies – by those who follow the Nolan principles – can be put in a test harness to figure out what the principal components are, and what they actually do, and be analysed using representative populations that take account of protected characteristics and test whether the

---

[7] As the last resort, courts are the final recourse of law for private life.

[8] "For example, we will do more to collect data on the protected characteristics of those who use the courts and tr bunals in a way that will make it far easier to identify and tackle disproportionalities."- paragraph 6,Ministry of Justice, Evaluating our reforms, Response to PAC recommendation 4, January 2019. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/775588/Public_Accounts_Committee_Recommendation_4_31_Jan_2019pdf.pdf

[9] https://www.gov.uk/government/news/hmcts-announces-expert-advisor-on-open-data-and-academic-engagement

[10] Byrom, N (2019) "Developing the detail: Evaluating the Impact of Court Reform in England and Wales on Access to Justice" https://research.thelegaleducationfoundation.org/research-learning/funded-research/developing-the-detail-evaluating-the-impact-of-court-reform-in-england-and-wales-on-access-to-justice

algorithm is the implementation of the integrity, honesty, and decision making (leadership) that the Nolan principles require.

New standards must be grounded in the long traditions of academic and practical work that underpins principled frameworks and practice. Standards applied to explaining decision making in automated systems should be derived from the Bingham Principles directly, and for policy processes from the Nolan Standards – especially around objectivity, fairness, and transparency – rather than be retrofitted off current human practice. Human practice being an institutional norm which assumes that the status quo is not worth the effort to improve it, which has always been the argument of the status quo, and not that of innovators. It is a sign of the political seductiveness of these new technologies and their corporate advocates that the narrative has changed, embodying the view that the powerful should never be challenged with principles, but should be allowed to do what is easiest for them. The 'ethics' frameworks are a cargo cult of industry where profit and contracts are paramount, externalities are ignored, and caveat emptor.

Rather than algorithms and AI merely reflecting current practice, all uses should show how they comply with the Committee's 7 Principles of Public Life, and the Bingham Principles on the Rule of Law (for decision making).

medConfidential
May 2019

Annex: How the Standards of Public Life apply to algorithms, data, and AI

Under the current Terms of Reference, the Committee "can examine issues relating to the ethical standards of the delivery of public services by private and voluntary sector organisations, paid for by public funds, even where those delivering the services have not been appointed or elected to public office."

Public standards apply to actions in public life and, just like obligations under the rule of law, they apply *however* decisions are made.

Announcing what became the Committee, then Prime Minister John Major said:[11]

> "If the rules governing conduct in public life are vague or unsatisfactory, Nolan will clarify them. But his task is not just to meet immediate questions. It is to act as a running authority of reference – almost, you might say, an ethical workshop called in to do running repairs."

And paragraph 6 of the Committee's First Report says:[12]

> "Frequently in our work we heard the expression 'grey area' used as a rationalisation of morally dubious behaviour. The ubiquity of the phrase, and the implication that some no

[11] Speech at the Lord Mayor's Banquet, 14 November 1994. Quoted: in CSPL "The First Seven Reports - a Review of Progress" https://webarchive.nationalarchives.gov.uk/20131003071401/http://www.public-standards.gov.uk/wp-content/uploads/2012/11/First7Reports_ProgressReview.pdf
[12] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/336919/1stInquiryReport.pdf

longer seem to be certain of the difference between what is right and what is wrong in public life, concern us. When people in public life are in doubt whether a particular action is consistent with the standards expected of them, the only proper course is not to do it."

Somewhat prophetically, paragraph 61 notes:

"Commercial organisations which have gone through what is called 'de-layering' have recognised that increased management responsibility at lower levels may confront junior staff with ethical issues of which they have had no previous experience. They may need support which is no longer provided by the line management hierarchy. The civil service is increasingly in the same position and at the same time is being asked to become more flexible and entrepreneurial in its provision of services."

SUBMISSION 6

Written by Crofton Black and Cansu Safak
Published: May 8, 2019

This report gives findings from a six-month research project by the Bureau of Investigative Journalism looking at the purchase of data systems and algorithmic processes in the public sector. The Bureau is a non-profit organisation that pursues investigative journalism in the public interest.

Thanks to Emma Prest, Jessica Purkiss, Josh Stein, Emilia Terracciano
Report designed by Rob Minto
Supported by a grant from the Foundation Open Society Institute in cooperation with the Information Program of the Open Society Foundations

Illustration: Chris Campbell
Images: Shutterstock (735439414, 374692213, 125192411)
Home Office, Marsham Street: Steve Cadman, Creative Commons
Graphic: Crofton Black / Maltego

Crofton Black is a researcher, journalist and writer with extensive experience of complex investigations into counter-terrorism, military contracting, social media and propaganda.

Cansu Safak is a researcher with a background in the examination of predictive analytics in the public sector and complex FOI projects.

THE BUREAU OF INVESTIGATIVE JOURNALISM

Summary

- Development of algorithmic and data-driven systems is frequently predicated on austerity - doing more with less.
- The adoption of such systems, the combining of legacy databases and the roll out of “digital by default” services is a major driving force in public policy.
- Information technology contracting for government is a flourishing area: although still dominated by traditional big names, many lesser known companies are also offering a wide variety of services.
- The Home Office’s ambitious plans for new data-driven systems have required the assistance of over 40 companies in just the last two years.
- We traced several types of government purchasing activities for digital systems, finding over 1,800 companies selling simpler or shorter-term consultancy, software and storage services and almost 300 providing more complex requirements.
- Many authorities were unwilling or unable to specify how and why they purchased these services, however, or what their precise specifications were.
- Public authorities - national and local - are supposed to keep transparent and accessible records of the services they purchase (in part to comply with the Public Contracts Regulations 2015). We found that this was rarely the case.
- Government transparency datasets are an inadequate tool for understanding purchases, particularly in the case of highly diverse large companies which offer a multiplicity of services (true of many major government contractors). The UK lags massively behind the US’s granular approach to public spending available through the Federal Procurement Data System, for example.
- Transparency - and therefore accountability - over the way in which public money is spent remains a very grey area in the UK. This is concerning, particularly at a time when the state is driving a complex data-driven revolution predicated on saving money through major digital transformation programmes and legacy system overhauls.

Introduction

AI, algorithms, deep learning, big data - barely a week goes by without a new revelation about our increasingly digital future. Computers will cure cancer, make us richer, prevent crime, decide who gets into the country, determine access to services, map our daily movements, take our jobs away and incarcerate us. Successive innovations spark celebration and concern. The UK should be "ready, willing and able" to profit from the economic potential of being a world leader in the artificial intelligence industry, a recent government report emphasises.[1] Academics and civil society, meanwhile, sound warnings over corporate accountability, the intrusiveness of personal data and the ability of legal frameworks to keep pace with technological challenges.

At conferences, during interviews, or over a pint in a pub, we've consistently heard one refrain: people are convinced that the growth of technology in the public sector has hugely important ramifications, but are baffled as to what exactly is going on and who is doing it. This report is a first step in remedying this. It gives a summary of some key findings from a six month scoping project into the use of algorithmic, data and digital systems in the UK public sector. We focus on how these systems are purchased, who from and who by, and we offer insights and tools for others to build further investigations on.

Our scoping exercise coincided with a number of events relating to the examination of government data systems. Two major reports were published: *Data Scores as Governance: Investigating uses of citizen scoring in public services* (Data Justice Lab, Cardiff University) and *Automating Society: Taking Stock of Automated Decision-Making in the EU* (AlgorithmWatch).[2] The UN Special Rapporteur on Extreme Poverty and Human Rights, Philip Alston, concluded a two-week mission to the UK with a scathing assessment of the implementation of the "digital by default" Universal Credit system.[3] A major cross-disciplinary workshop on "Algorithms and Society" was held at the Vrije Universiteit, Brussels.[4] The NGO StopWatch organised an event in London, hosted by Amnesty International, scrutinising the London Gangs Matrix, while Liberty released *Policing by Machine*, a round-up of public information on new policing technologies.[5] All these events attest to significant, and growing, interest in and concern for the role that data generated by or about citizens is playing in government systems which affect them. Among very many stakeholders, there is a strong feeling that the time has come for an urgent and wide-ranging debate.

Questions abound over how the debate is framed. What do we mean by an algorithmic or automated system? What is the difference between how such systems are designed and

[1] https://pub ications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf
[2] https://datajusticelab.org/data-scores-as-governance/ and https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Pub ikationen/GrauePub ikationen/001-148_AW_EU-ADMreport_2801_2.pdf
[3] https://www.ohchr.org/documents/issues/poverty/eom_gb_16nov2018.pdf
[4] http://www.privacysalon.org/programme
[5] http://www.stop-watch.org/events/details/report-launch-being-matrixed-the-overpo icing-of-gang-suspects-in-london and https://www.libertyhumanrights.org.uk/policy/report-po icing-machine

how they are used in practice, and how can each of these be evaluated? At what point does an advisory tool become unchallengeable revelation (or mysterious cargo cult)? Which potential issues lie in back-end processing and which in front-facing design and implementation? Are more problems caused by combining datasets or by failing to combine them? To what extent are the negative effects associated with new technologies actually indicative of persistent structural social ills? Are new approaches formalising inequalities that previously were less clearly defined and entrenched? Do data-driven systems perpetuate and codify unconscious biases, or do they offer a means for correcting formerly unarticulated prejudices?

To guide our work, we decided to focus in particular on the interface between public sector buyers of data systems and private sector sellers. Past experience has shown us that government procurement, and the traces it leaves in public (or at least potentially public) data can usefully be leveraged as a starting point for understanding diverse, and sometimes hidden, state activities.[6] The strategy through which a service is procured can illuminate the context and thinking behind the project, while documentary traces left by the process can point to concrete and specific details which are often obfuscated in official narratives.

Media, NGOs and academics in the UK have illuminated a number of private-public technology initiatives, but it has remained hard to assess the landscape in general. Conversations with sources suggested to us that this wasn't simply a shortcoming of our own perspective, but that, even within government, knowledge of what is going on remains fragmentary.

What we offer here is not a map, which would be a massive undertaking, but rather some practical methods which interested parties can use as starting-points for investigating their own areas of focus.

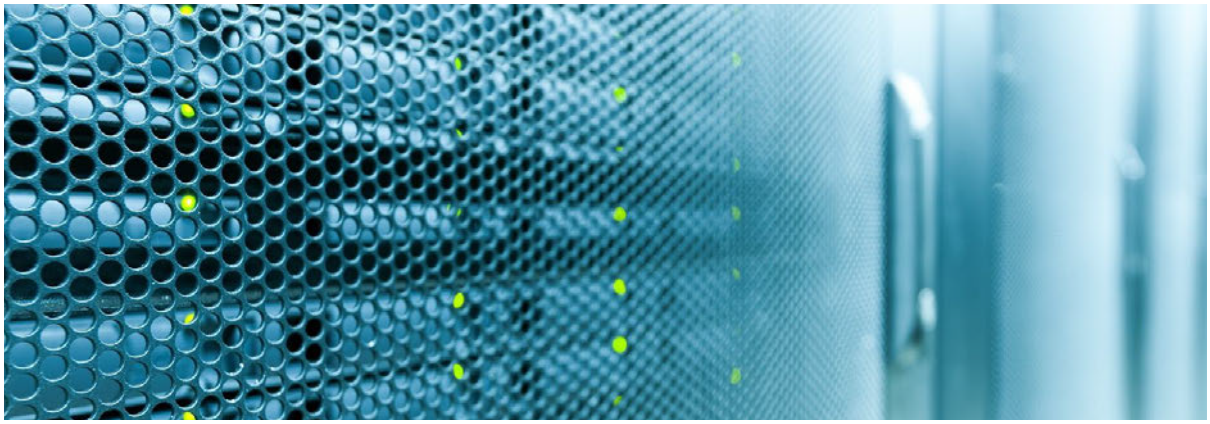
In the course of preparing this report we carried out dozens of interviews with experts and insiders. We are not publishing details of these interviews, but they have informed our approach and we are grateful to all those who agreed to talk to us. We have been particularly assisted by Swee Leng Harris, Tom Longley, Amber Marks and Sam Smith.

This report was researched and written by Crofton Black and Cansu Safak, with valuable contributions from Emma Prest. Technical assistance was provided by Charles Boutaud. We used WhatDoTheyKnow Pro to manage our FOI requests.[7] The project was kindly supported by a grant from the Foundation Open Society Institute in cooperation with the Information Program of the Open Society Foundations. We are grateful to Becky Hogge for actively encouraging this project.

[6] See, for example, The Bureau's work on drones and image screening (https://www.thebureauinvestigates.com/stories/2015-07-30/reaping-the-rewards-how-private-sector-is-cashing-in-on-pentagons-insatiable-demand-for-drone-war-intelligence) and propaganda in the Iraq war (https://www.thebureauinvestigates.com/stories/2016-10-02/fake-news-and-false-flags-how-the-pentagon-paid-a-british-pr-firm-500m-for-top-secret-iraq-propaganda).
[7] https://www.whatdotheyknow.com/

Investigating Procurement

Spending taxpayers' money comes with strings attached - including the need, at least in theory, to be somewhat transparent about where the money goes and why it goes there. As a result of this, a focus on government procurement can offer useful insights into state activities.



We split our investigation into two phases. First of all we looked at what existing transparency data could tell us about UK public sector investments in IT projects of potential interest. We used various portals and datasets, which we discuss below. Our investigation was carried out partly through automated techniques, including building scrapers to extract information from websites. We give a snapshot of our results, relating to contracts with the Home Office, in the appendix.

We then filed a group of Freedom of Information requests to test how public authorities maintained records of what they had purchased and why, using the data we had gathered in the first phase as source material.

We found that hundreds of companies are selling digital services to the government, but that the mechanisms which supposedly keep track of what services are being sold are far from effective. This is in striking contrast to the government's public commitment to open and transparent data.

Exploring Transparency Data

Public authorities in the UK make some information on their past and future purchases available for download and analysis. The UK likes to portray itself as a model for open data

standards.[8] We tested various portals used by the government to see which ones offered a useful insight.

Digital Marketplace

The Digital Marketplace is a gov.uk search tool which helps purchasers "find technology or people for digital projects in the public sector."[9] It offers streamlined procurement of digital services ("outcomes") and individuals ("specialists"), as well as a menu of pre-existing cloud services (hosting, software and support) and data storage.[10]

Digital Outcomes and Specialists (DOS) Framework

Registered users can create advertisements with written requirements for a particular "outcome" using the Digital Outcomes and Specialists framework. They are supposed to follow a series of steps: locating a list of suppliers, budgeting, writing and publishing requirements and evaluation criteria, answering supplier questions, evaluating and shortlisting supplier applications and finally awarding a contract.[11] The process is meant to take around four weeks.[12]

Suppliers (and members of the public) can access the advertisements posted using this framework via a keyword search box, some simple filters and a .csv download.[13] The search results as displayed on the website give project title, buyer, opportunity dates and a summary paragraph about what the project is intended to do. Clicking on the title leads to a new page with a more detailed description following a consistent structure. Each advertisement has an ID number, which is used to construct its URL ("https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/1111").

The .csv download offers metadata for opportunities (ID number, title, buyer, category, dates, link to advertisement etc.) but does not include the summary paragraph which explains what the work will do.

The government publishes "statistical data sets" of DOS spending.[14] These provide aggregated monthly figures, since June 2016, of how much each supplier has been paid by each authority, but do not separate spending out into projects. A third-party website,

---

[8] See, e.g., https://www.gov.uk/government/pub ications/uk-digital-strategy/7-data-unlocking-the-power-of-data-in-the-uk-economy-and-improving-pub ic-confidence-in-its-use: "The true potential of data can only be harnessed if it is open for use by others. The UK leads the world in open data, and the government is committed to building on this and being open by default."

[9] https://www.digitalmarketplace.service.gov.uk/

[10] https://www.gov.uk/guidance/digital-outcomes-and-specialists-buyers-guide

[11] https://www.digitalmarketplace.service.gov.uk/buyers/frameworks/digital-outcomes-and-specialists-3/requirements/digital-outcomes

[12] As our findings below show, buyers sometimes find even this quite short process too arduous.

[13] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specia ists/opportunities

[14] https://www.gov.uk/government/collections/digital-marketplace-sales#digital-outcomes-and-specialists-sales

govspend.org.uk, offers a .csv download of DOS opportunities with the convenient addition of the contract amounts, where these have been published on the DOS website.[15]

We built a scraping tool to extract all the summary paragraphs from the DOS website. We combined these with the .csv download of opportunities and with the contract values derived from Govspend. This meant that we could filter opportunities by keywords appearing in the summary paragraphs, to give a more useful insight into past opportunities.

We also scraped the full webpages for each individual opportunity as separate .txt files, allowing keyword search across full tender descriptions.[16]

In November 2018, using this method, we created a dataset of 2,389 opportunities, ranging in date from 28 April 2016 to 6 November 2018. (By 3 April 2019 there were 2,754 opportunities listed on the website.)

This method offers a useful snapshot of services that public authorities are looking for, and a more limited insight into which companies were awarded DOS contracts. 1,613 of the 2,389 opportunities in our combined dataset do not indicate which company received the contract award. In most of these cases, the website states merely that applications are closed and gives the number of companies which completed applications (but not their names). In some cases the website states that a contract was withdrawn, cancelled, or no suitable suppliers responded. 776 tenders in our dataset do state which company won the award. These 776 awards went to 292 companies. Those which are recorded as having received more than 5 awards are:[17]

Methods Business and Digital Technology Limited	24 awards
Interact Consulting Limited	18
Kainos Software Ltd	16
People Source Consulting Ltd	15
FUTUREGOV. LTD	14
Keystream Healthcare Resources Limited	14
PA Consulting Services Limited (UK)	13
LA International Computer Consultants Limited	13
Deloitte LLP	11
Engine Partners UK LLP (The Engine Group)	11

[15] https://www.govspend.org.uk/dos.php
[16] We found Sublime Text to be a useful tool for searching across files: https://www.sub imetext.com/
[17] Company names are drawn from the data as is, without attempting to adjust for consistency.

Lucid Support Services Ltd	11
The Dextrous Web (trading as dxw digital)	10
Made Tech Ltd	9
John White PM Ltd	9
Malikshaw Limited	9
Capgemini UK plc	8
Softwire Technology Limited	8
IBM United Kingdom Ltd	7
Valtech Ltd.	7
Redweb Ltd	7
Mercator IT Solutions	6
Agilesphere LLP	6
BAE Systems Applied Intelligence Limited	6
Rainmaker Solutions	6
CGI	6
COMXPS Limited	6
Unboxed Consulting Ltd	6
Lagom Strategy Ltd	6

G-Cloud Framework

The Digital Marketplace also offers a large number of off-the-shelf purchases. In such cases, suppliers list pre-existing software or services for the buyer to choose from, rather than the buyer writing an advertisement describing a bespoke service. These off-the-shelf services can be bought via the G-Cloud framework.[18] They are divided into three categories: cloud hosting, cloud software and cloud support, and are simpler to purchase than Digital Outcomes because the service descriptions already exist: buyers just need to make a search for the type of service they require, assess the results and award a contract to whichever supplier they deem appropriate.[19]

[18] https://www.gov.uk/guidance/the-g-cloud-framework-on-the-digital-marketplace
[19] https://www.gov.uk/guidance/g-cloud-buyers-guide

As with the DOS framework, the government publishes a statistical dataset of G-Cloud spending, which gives aggregated values for spending by supplier and purchasing department per month.[20] Since some suppliers offer dozens or even hundreds of services on G-Cloud this is not helpful for determining what the government has actually bought.

We scraped the G-Cloud search website to give us a .csv file of nearly 25,000 services offered by over 3,000 suppliers. As with the DOS framework, our scraped data included summary paragraphs and links to full descriptions of each service, but not the complete text of each full description.

The data highlights a significant number of companies offering IT services within areas of particular interest.[21] We located

- 88 companies offering 101 separate “machine learning” services
- 73 companies offering 89 separate “big data” services
- 47 companies offering 50 separate “predictive” services
- 41 companies offering 56 separate “artificial intelligence” services
- 17 companies offering 19 separate “biometric” services
- 8 companies offering 6 separate services relating to facial recognition and expression analysis

Some of these companies are well-known government suppliers. Many are not.

The public sector as a whole spent over £3bn on G-Cloud purchases in 2016-18. The top ten buyers over this period were:

Home Office	£446,575,353
HM Revenue & Customs (HMRC)	£229,396,059
Ministry of Justice	£213,446,086
Department for Work and Pensions	£201,058,135
Cabinet Office	£87,981,583
Department for Education	£87,211,839
Student Loans Company	£74,453,356

[20] https://www.gov.uk/government/collections/digital-marketplace-sales#digital-outcomes-and-specialists-sales
[21] As discussed below, shortcomings in the transparency data mean that without (and sometimes notwithstanding) further investigation it is impossible to know which of these services have been purchased.

Driver and Vehicle Standards Agency	£57,241,306
Driver and Vehicle Licensing Agency	£56,953,412
Office for National Statistics	£48,847,066

Other Tendering Data

The Digital Marketplace offers a relatively streamlined mechanism for government departments to buy IT products and services, but it is only one of many existing portals for government contracting work. All projects on the Digital Marketplace are digital, but not all digital projects are to be found on the Digital Marketplace.

Spend Network

To assist in our investigation we were kindly provided with access to the tender database collated by Spend Network.[22] This offers some 270,000 tenders from across local and national government, from 2016 onwards.

We built an API query to extract tendering documents containing specific keywords. On an initial test run we extracted documents containing 11 basic key terms (algorithm, automated, big data, decision making, decision-making, match, predictive, profile, risk based, risk-based, regression). This produced 3,874 tenders over 61 domains. The most frequently occurring domains for this search were

- www.contractsfinder.service.gov.uk 1,484
- www.sell2wales.gov.wales 593
- www.digitalmarketplace.service.gov.uk 467
- in-tendhost.co.uk 318
- procontract.due-north.com 243
- irl.eu-supply.com 181
- www.publiccontractsscotland.gov.uk 102

12% of all the tenders located appeared on the Digital Marketplace. 38% came from the Contracts Finder website.

[22] https://spendnetwork.com. We are grateful to Ian Makgill of Spend Network and George Brown of Open Opps for facilitating access for us.

We subsequently ran a much broader extraction with a list of 59 terms, focusing not only on types of technology or methodology (as in the first list) but also on service areas in which IT systems are being used (e.g. biometric, housing, fraud, arrears). This provided a much larger dataset of over 30,000 tenders from over 100 domains.[23] The top 10 domains by frequency in this list were

- www.contractsfinder.service.gov.uk 13,571
- www.sell2wales.gov.wales 5,603
- procontract.due-north.com 2,099
- irl.eu-supply.com 1,984
- www.digitalmarketplace.service.gov.uk 1,548
- in-tendhost.co.uk 1,480
- www.publiccontractsscotland.gov.uk 1,105
- www.competefor.com 453
- www.defenceonline.co.uk 265
- etendersni.gov.uk 231

Digital Marketplace accounted for a much lower percentage (5%) than on the previous extraction - unsurprisingly, since the search terms were much less narrowly focused on digital processes - while Contracts Finder slightly increased its share (44%).

Contracts Finder

Our analysis of Spend Network tenders indicated that the dominant data source for the type of information we were looking for was Contracts Finder, a search engine for public sector contracts run by the Crown Commercial Service (CCS).[24] At time of writing, Contracts Finder contains a database of over 159,000 open, closed and awarded government tenders, accessible via a search bar and a series of filters. Some tenders have linked contracting documents. The site also offers a .csv download.

Unfortunately, Contracts Finder's search and download features make it difficult to use the site to perform the type of analysis we required.[25] The download facility is limited to 1,000

[23] We excluded results from TED (the EU's Tenders Electronic Daily), although some of these would have related to UK purchases. The resulting dataset included 5,818 tenders dated 2016, 10,904 dated 2017, 11,551 dated 2018 and 2,658 dated 2019.
[24] CCS is described as an executive agency "sponsored by the Cabinet Office" and "providing commercial services to the pub ic sector": https://www.gov.uk/government/organisations/crown-commercial-service.
[25] Without going into exhaustive detail, the search bar struggles to parse boolean searches alongside precise word formulations.

rows. We requested access to the full dataset in .csv format from CCS at the start of February 2019, but had not received a substantive response by the end of April. Daily .csv downloads of Contracts Finder data are available, but hard to use because they do not consistently maintain the same column configuration and therefore cannot be easily combined with each other.[26] Furthermore, there is no simple way of correlating entities in the downloaded datasets with the website search results.[27]

Third-Party Datasets

A number of free or commercial websites offer access to data drawn from Contracts Finder and other government portals, some with better accessibility than Contracts Finder itself. We found Bidstats to be a useful resource for keeping up with new tender opportunities.[28] Bidstats combines data from Contracts Finder and the Official Journal of the European Union (OJEU). The tenders are organised by sector, but can also be filtered according to company, government department, sector or a wide variety of hyperlinked tags.

As noted above, we incorporated data from Govspend,[29] a website offering analysis of transparency data relating to the Digital Marketplace. Unlike the aggregated transparency data released by the government, Govspend in some cases includes individual contract amounts.

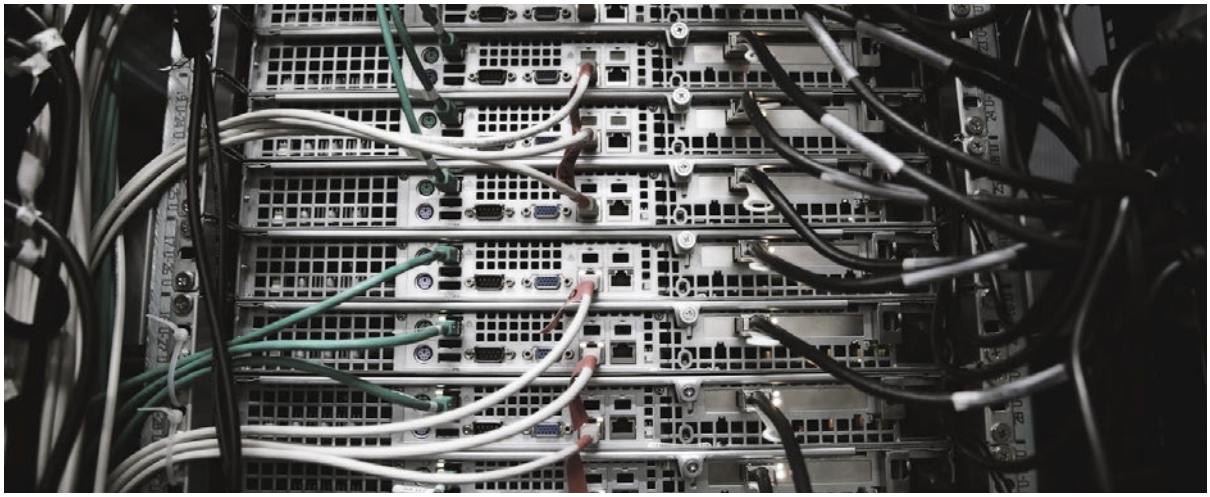
Another such initiative is Appgov,[30] which offers search functionality on collated spend data and a list of major central government projects with links to annual reports associated with them.

Audit Trails

Principles

When civil servants buy services from the Digital Marketplace, they are supposed to maintain accessible audit trails of how they made those purchases. The G-Cloud buyer's guide states that "if you award a contract through the G-Cloud framework, you must be able to show that your assessment of services was fair and transparent."[31] As a G-Cloud buyer, "you should keep your own record of your communications with suppliers, including any early market engagement, clarification questions, emails and face to face conversations you have." The guide's recommendations extend to including in the audit trail the search terms used when looking for services.

[26] https://data.gov.uk/search?filters%5Bpublisher%5D=Crown+Commercial+Service
[27] The .csv has a "notice identifier" but this does not correspond to the URL for the tender and needs to be searched for in a two-stage process in the search engine.
[28] http://bidstats.uk/
[29] https://www.govspend.org.uk
[30] https://www.appgov.org/
[31] https://www.gov.uk/guidance/g-cloud-buyers-guide



Likewise, the Digital Outcomes and Specialists Audit Trail Guidance stipulates that buyers need "a record of all the activities and decisions you make at every stage of the buying process ordered by date."[32] This is necessary to ensure compliance with the Public Contracts Regulations 2015.[33]

The Public Contracts Regulations are a statutory instrument which require, in part, that "contracting authorities shall document the progress of all procurement procedures, whether or not they are conducted by electronic means," and that "contracting authorities shall ensure that they keep sufficient documentation to justify decisions taken in all stages of the procurement procedure," including "communications with economic operators and internal deliberations", "preparation of the procurement documents", "dialogue or negotiation" and "selection and award".

DOS publishes clear and simple principles underlying audit trail maintenance, stating:

"Your audit trail must:

- be kept for at least 3 years
- be stored in an accessible format, in a place that other people in your team have access to, for example a shared folder
- keep a record of the buying process from your first contact with suppliers, for example from early market engagement until the contract (or 'call-off') ends"[34]

We found, however, that many departments appear not to follow these principles, and that as a result public understanding of what services the government is buying, and how it is buying them, is greatly curtailed.

[32] https://www.gov.uk/guidance/digital-outcomes-and-specialists-audit-trail-guidance
[33] http://www.legislation.gov.uk/uksi/2015/102/contents/made
[34] https://www.gov.uk/guidance/digital-outcomes-and-specialists-audit-trail-guidance

Findings

Only a quarter of authorities we asked were willing or able to provide some form of audit trail for their G-Cloud purchases.

In January we sent requests to 40 authorities asking for titles, ID numbers and audit trails of G-Cloud purchases from specific companies. We knew, from our data analysis, that these authorities had purchased some services from these companies under the G-Cloud framework, but owing to the shortcomings of the transparency data, outlined above, we did not know which services they had bought.

By the end of April, 35 of these authorities had replied.[35] 13 of them refused the request entirely, while a further 5 provided some service titles but no further information. In most of these cases, the authorities stated that the request exceeded their cost thresholds because, contrary to the guidance, they did not maintain their audit trails either centrally or accessibly.

One authority - Central Bedfordshire Council - claimed to be unable to provide even a list of services purchased because "the Officer responsible for Procurement activities is no longer with the Organisation [and] information has not been transferred to Procurement system." Cambridge Constabulary stated that "we do not have access to an audit trail." The Defence Infrastructure Organisation - part of the Ministry of Defence responsible for training facilities - replied that it would take them at least 48 days to locate and extract audit trails on purchases from two companies between 2015 and 2019.

Aylesbury Vale District Council explained that "The information related to G-Cloud contracts were kept on specific projects filing structures; and not necessarily available for projects and contracts that were concluded and/or terminated. … It is unknown the number of documents to review, and the document management structure changed in the recent years. There would be a requirement to dedicate an officer or two engaging with many departmental managers, project managers and possibly former employees to fully respond to this request."

The Serious Fraud Office stated that "our initial search has made it apparent that it would not be possible to respond to your request without referring back to archived backup tapes. The data requested is not actually stored in a central location and in order to find it we would have to conduct an extensive search across these backup tapes to find and collate it. So far we have already spent over seven hours of work determining exactly what data is held. We estimate that to recall and restore the backup tapes would require in excess of 18 hours of work. Furthermore, if the tapes were fully restored, it would take considerable staff time to conduct searches across the material and to identify any relevant information."

Two authorities argued that they could not release their audit trails because of commercial sensitivity. Two authorities declined to disclose the services they had purchased on the

[35] Requests were made on 24 January or before, with the statutory reply period expiring by late February. One authority - the National Crime Agency - anomalously falls outside the FOI law.

basis that the information was already publicly available on the Contracts Finder website; we did not find this to be reliably the case, however.[36]

In four instances, authorities claimed that they had not made the specified G-Cloud purchases, even though we had made our requests on the basis of G-Cloud spending data. It is unclear how to account for this.[37] Two authorities stated that purchases attributed to them in the spending data had actually been made by other authorities; in one instance, the other authority then denied that it had made any purchases.

Some authorities were seemingly able only to identify a fraction of the services that G-Cloud data recorded them as having bought. The government's statistical dataset states that in 2016-8 Transport for London bought G-Cloud services from BAE Systems Applied Intelligence, IBM, Deloitte and BravoSolution (among others), whereas TfL claimed only to have located one purchase out of this group. HMRC initially declined to provide a list of services purchased from five companies. Following a request for internal review, they listed services bought from three of the five companies. No explanation was given as to whether or why they had failed to locate services from the other two companies.

One authority - Staffordshire Police - did not technically refuse the request, but responded without giving any of the information requested, stating merely that the information we asked for was part of a broader contract with Boeing Defence UK.

Five authorities provided only the titles of the services they had purchased.[38] Nine (23% of those originally asked) provided some form of audit trail in response to our first request, ranging from very sparse to quite detailed.

At time of writing, five authorities (Peterborough City Council, Islington Council, the Metropolitan Police, City of London Police and Nottinghamshire Police) had failed to finalise a response, more than three months after the request was first made. The Met said they were working on it but that "the information is currently 'dispersed' and it is taking time to find 'what we hold, and to collate it'."

We also made some requests for purchases under the Digital Outcomes and Specialists framework. In response to our request for audit trails of three purchases, the Ministry of Justice stated that "we believe that the cost of extracting the information would require manually going through [an] estimate of 84 contracts files (manual & electronic) … This information is not held centrally and the cost to obtain some of the information from various units and directorates within the department would exceed the appropriate limit." In response to a request for internal review, the MOJ upheld its refusal: "The MoJ stores all commercial documents on one central and accessible system; however, the full audit trail

[36] The Cabinet Office asserted that "details of call-off contracts are pub ished on Contracts Finder." A search for "North Highland" on Contracts Finder produced one contract between GDS and that company, awarded in March 2015, but this did not appear to match the series of expenditure records recorded under G-Cloud between January 2015 and December 2018.
[37] The Digital Marketplace website states CCS's transparency figures for G-Cloud are provided monthly by suppliers: https://www.gov.uk/government/collections/digital-marketplace-sale.
[38] Some authorities could not locate the G-Cloud ID numbers for the services they had purchased.

(including communication with suppliers, notes, etc.) are not required to be stored in such a centralised way. Much of the key information is saved centrally, while other areas of the audit trail may be saved within individual and/or local drives, meaning that to locate, retrieve and extract this disparate information would exceed the cost limit."

The Home Office declined a request for audit trails of nine different procurements on grounds of cost (providing an estimate that it would take 72 hours to retrieve information on these procurements). A narrowed request was then refused on grounds of commercial interest: "Release of the withheld information would provide competitors with information, not available to them by any other means, about current service providers. This would create an unfair advantage resulting in a prejudice to the commercial interests of the company or companies concerned. Disclosure would also prejudice the Home Office's commercial interests by damaging commercial relationships with contractors and service providers."

Not all departments believe that disclosing audit trail information is detrimental to commercial interests, however. The Foreign and Commonwealth Office had refused our first request on grounds of cost, but responded positively to a narrower request. The disclosures totalled hundreds of pages of emails, evaluations and contractual material. Key data was still missing. Nonetheless, the FCO's response shows that disclosing detailed audit trail information is possible.

Case Study: G-Cloud Searching and Evaluation - Video Justice

In a few cases we received enough information to draw some conclusions about how authorities locate and evaluate G-Cloud services. Here is one example, chosen because it gives a good insight into a relatively simple search and evaluation process, and because its topic - digital justice provision in an age of austerity - is of considerable social importance.

Starting in 2015, Sussex Police procured various services via G-Cloud in relation to the development of a "video enabled justice system", in partnership with other police forces and national departments. The overriding goal was "delivering justice in an environment where there are fewer court buildings and less staff", through the use of "video by default". Sussex Police considered a number of avenues for procuring projects associated with this system. In one such instance, "consultancy services for the provision of a target operating model", in-house management was discounted because expertise and capacity were unavailable. The Home Office's Bluelight framework, commonly used by police forces, was discounted because of "the perception of a lack of independence and impartiality" and "connotations of it being a police-led project when it is essential that partners feel they are operating in an environment of inclusion." Use of an existing partnership with Deloitte would have created conflicts with another programme, while issuing a tender invitation was unattractive because of time scales (which were described as "extremely tight"). The G-Cloud framework was preferred because "it is a service for short term contracts [and] it is more efficient for selecting a suitable supplier through a direct-award process rather than sending out an [invitation to tender]." A similar process was followed for the subsequent "detailed business case" the following year. Time was too tight to use the DOS framework because of the

tendering invitation requirement, and other frameworks which, like G-Cloud, lacked tendering invitation requirements were seen as less specialised towards digital acquisitions or more limited in the range of suppliers offered.

Having decided to use the G-Cloud framework, procurement officers set about searching for the company which would offer consultancy on the “Video Enabled Justice system - Phase 2 Detailed Business Case” by inputting a wide variety of terms in the G-Cloud search box. These search terms included “connects video”, “joint working”, “end to end”, “reduced business travel”, “return on investment”, “staff satisfaction” and “digital transformation”, among others. Many of these terms returned dozens or hundreds of results; these were filtered and limited by the addition of the term “video” to each. An intermediary longlist produced a variety of services from Accenture UK, Affinity Digital Technology, AVM Impact, Dimension Data, Cloud Technology Solutions, Modality Systems, Involve Visual Collaboration and numerous other companies. Accenture, Modality Systems and Involve made it to the short list, where they were scored on service definition (via subcategories of relevant and specific experience, method and approach and mobilisation speed) and pricing, with Accenture coming top.

Perspectives

As well as examining procurement processes and records, we carried out dozens of interviews with experts, insiders and members of civil society. The purpose of these interviews was to help us situate our findings about the buying and selling of IT and data systems in the general context of current policy initiatives and debates.

Several interviewees suggested a strong link between the adoption of data-driven systems and the programme, dominant in recent years, of “austerity”. Documents produced by both sellers and buyers frequently reference the need to do more with less, and systems are sold to the public sector with the promise that they will offer savings or be more efficient than existing methods. Questions exist as to how direct the link between buying IT systems and saving money is, however. There is evidence that automation of processes saves time. But more than one interviewee told us that while a good IT system costs a lot, the only real way for public authorities to save money is to cut down on staff or property.

Several interviewees expressed frustration with current narratives about algorithmic systems. Sellers, buyers and the media have all been known to exaggerate the sophistication of systems and to package traditional methods of statistical modelling as being “artificial intelligence”.

Many interviewees were not entirely negative about authorities trying to do a better job with more data. But more than one cautioned that data-driven systems need to be contestable, and that this requires them to be transparent.

Several interviewees emphasised that, looking at the matter from the other end, the design and implementation of systems need to be properly in line with the needs of users - typically officials and frontline staff - as well as the ideas of their managers. Otherwise they risk being a waste of money or producing answers to the wrong questions.

A narrow focus on algorithms or data usage can be counterproductive when trying to assess complex systems. Interviewees who discussed the Universal Credit system with us saw its problems as being linked to poor communication, poor training, poor advice and poor resourcing more than to the algorithms embedded within the system. Others highlighted the gap between what a system is built to do and the culture of how it is used, and how this can change over time. Decisions made or suggested by an IT system are a subset of a broader range of decisions made by human users. The function of the IT system can’t be assessed without evaluating the overall context of the management and working culture within which its human users operate.

Conclusions

As we stated at the outset, a recurring refrain throughout our interviews was the basic dearth of structured information about what public authorities are doing with their IT investments. This report aims to show some ways in which this problem can be addressed, while pointing to significant barriers that investigators still need to overcome.

We've examined some of the avenues through which the government buys IT services, identified companies (large and small) which are competing for public money and tested the transparency of the procurement process. We've also shown how existing open sources can be combined to form a more comprehensive picture of government purchases.

It is clear from our research that there is a flourishing ecosystem of companies offering digital, data and algorithmic services to public authorities. Over 1,800 companies are listed as having sold G-Cloud services (including software and hosting) to the public sector. 776 more complex awards via Digital Outcomes and Services went to almost 300 companies. As our appendix shows in detail, the Home Office's work over the last two years on several major transformational IT projects (spanning border, law enforcement and biometric databases) has required over 40 corporate partners operating under dozens of separate contracts.

In short, there is a very active industry competing for public money in exchange for data-driven and digital services. However, many authorities are unwilling or unable to specify how and why they purchased these services. This should ring alarm bells. The datasets which the government publishes in the name of transparency are inadequate,[39] particularly in the case of highly diverse large companies which offer a multiplicity of services - true of many major government contractors.

Transparency - and therefore accountability - over the way in which public money is spent remains a very grey area in the UK. This is concerning, particularly at a time when the state is in the midst of a complex data-driven revolution predicated on saving money through major digital transformation programmes and legacy system overhauls.

Such overhauls and transformations are likely to have a major effect on a significant proportion of the population - including the most vulnerable. Intentionally or not, they have the capacity to disrupt long-standing legal frameworks and policy norms around the way in which the state makes and implements decisions. Government contracting and outsourcing is traditionally an area which requires vigilance, especially regarding delivery of services and accountability. But without basic tools to assess what is going on and who is doing it, such vigilance is difficult.

We hope that our findings will prove helpful to the many people in journalism, academia, civil society, the legal profession and parliament who are currently grappling with these

[39] Compare the granular data on spending pub ished through the USA's Federal Procurement Data System.

significant developments in state bureaucracy and management. The potential for automated systems to offer “great possibilities for good and for evil” has been highlighted for decades.[40] Despite this, we still face an alarming deficit of knowledge about what these systems entail. Investigating the landscape of public-private partnership in the development of complex data systems offers some first steps to remedying this and provides a foundation for further inquiry.

[40] The classic formulation is that of robotics pioneer Norbert Wiener, *Cybernetics or control and communication in the animal and the machine*, New York 1948, pp. 37-9.

Case Study: The Home Office

Summary

The Home Office is one of the departments most active in the purchase of IT services. It is in the middle of a number of major projects which aim to replace expiring legacy systems and transform the management of immigration, visa and passport applications, border control and policing.[41] Its objective is to create integrated, "digital-by-default" systems and to become a "data driven organisation".

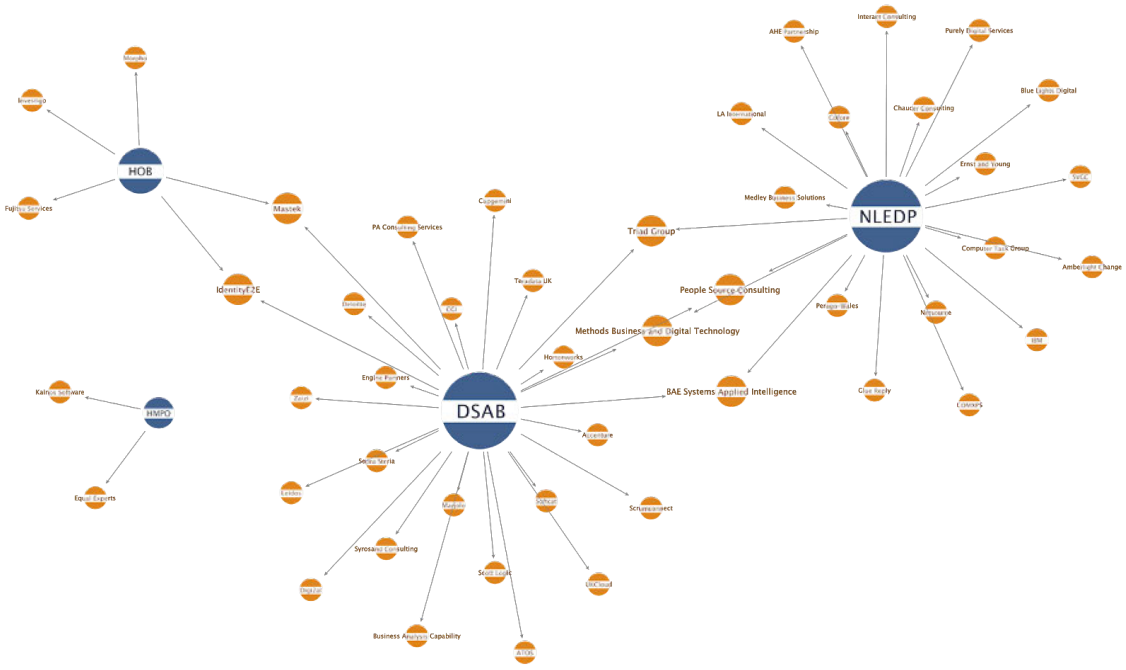
Searching through procurement data, we identified over 40 companies contracted to provide services or specialists to the Home Office since late 2016 on major data-driven transformation projects. We list 86 of these contracts below, grouped under four broad headings (Digital Services at the Border, the National Law Enforcement Data Programme, Home Office Biometrics and the Passport Office's Digital Application Processing).

Roughly 180 firms sold G-Cloud services to the Home Office over the same period. These are also listed below, although further research is required to identify what services were purchased in each case. Analysis of G-Cloud data shows that of the firms which were used by the Home Office, those with fewer services available offer biometric and document

[41] Some details of Home Office contracting in this area are discussed in Corporate Watch, *The UK Border Regime: A critical guide*, Oct. 2018, available at https://corporatewatch.org/new-book-the-uk-border-regime/.

authentication, machine learning and predictive analytics for fraud detection, risk scoring, and fusion of data from multiple sources, among other things.

An FOI request for audit trails of several purchases made by the Home Office via the Digital Marketplace was refused following a public interest test. The Home Office argued that disclosure would prejudice its commercial interest and that of the companies concerned.



Companies and Contracts

PA Consulting Services and **Accenture** are key partners in the Immigration Technology Portfolio. The tenders published for these services describe the goal as "automating existing manual, paper-based processes; and supporting decision-making to drive consistency."[42] Among the tasks are to "resolve the identities of individuals … manage information about individuals [and] perform and record the outcomes of checks, including calculations (eg. to assess whether a threshold in the Points Based System is met)." Questions therefore exist concerning to what extent the Home Office is moving towards a data-driven and/or automated risk-based approach in assessing immigration applications. The government is hoping to develop in-house capacity to run the Immigration Technology portfolio, but this is currently lacking: **Deloitte** has a £23.5m award to assist "transition … to the internal capability."[43]

"Digital Services at the Border" (DSAB) is the successor to the failed E-borders programme.[44] The Home Office wants to replace the two systems supporting E-borders (Semaphore and Warnings Index – 13 and 22 years old respectively) with the new Advanced Border Control and Border Crossing systems. DSAB also includes the Advanced Freight Targeting Capability (AFTC) system, replacing HMRC's Freight Tracking System, to identify cargo of interest.

BAE Systems Applied Intelligence has been awarded a £4.9m contract to provide the Borders and Immigration Advanced Border Crossing Delivery Team, "specifically focusing on the delivery of a digital capability to allow Border Force, Police and other key stakeholders to identify and assess threats who are about to travel to, and from, the UK."[45] The requirements include "large scale data ingest" and "information and analytics". **Triad Group** has been contracted to provide developers for the AFTC system "data ingestion/management project" and the HBASE big data platform,[46] while **Syrosand** has been contracted to "analyse the data for subjects (both people and goods) approaching the UK border and provide staff with analytics driven insights to allow intervention where necessary."[47] **Zaizi** is engaged in digitising processes for Border Force officers: in May 2018 the Home Office noted that "incumbent solutions are locally built solutions based on Excel and Access which do not scale to meet the present multi-user requirements or meet our ambition for becoming a more data driven organisation."[48]

The National Law Enforcement Data Programme (NLEDP) is being delivered by **IBM** (and other partners) and aims to upgrade legacy systems - notably the Police National Database (PND) and the Police National Computer (PNC) - into an integrated single-view system which

[42] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specia ists/opportunities/5768 and https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specia ists/opportunities/5771
[43] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specia ists/opportunities/5428
[44] https://www.nao.org.uk/wp-content/uploads/2017/10/The-UK-border.pdf
[45] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specia ists/opportunities/5403
[46] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specia ists/opportunities/2615 and https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specia ists/opportunities/2616
[47] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specia ists/opportunities/4908
[48] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specia ists/opportunities/6309

will link records of individuals from multiple sources (accessible to all law enforcement communities). It offers functions such as "analys[ing] data to identify links between people, objects, locations and events" and setting up "automated alerts for new or changed data and events of interest."[49] The **Police ICT Company** – currently owned by the Association of Police Authorities and Home Office – coordinates a variety of data activities and purchases across police forces.[50] It has "recently extended the national agreement with IBM providing access to analytics tools until 2020 … These capabilities extend beyond the i2 Analysis Software that analysts in every force are using already, to include predictive and video analytics tools."[51]

The Home Office is also in the process of developing the Home Office Biometrics Programme (HOB), to support the two legacy systems of IDENT1 and IABS (Immigration and Asylum Biometrics System) in a capability covering "3 primary biometric modalities of fingerprint, DNA and facial image associated with searches, verification checks and profile storage."[52] As stated in the Home Office Biometrics Strategy, "In future, HOB will provide a common facial matching service enabling the Home Office to realise efficiencies and ensure a more consistent approach to the testing, access controls and privacy controls associated with it. This will allow improvements in the technology and matching algorithms to enhance processes at Ports of Entry, Visa Application Centres and within passport applications."[53] **Fujitsu**, **Morpho** and **NEC** have been awarded contracts to deliver the Home Office Biometric Matcher Platform,[54] with **IdentityE2E**,[55] **Investigo**[56] and **Mastek**[57] providing associated services.

HMPO has announced a transformation agenda that describes its vision as shifting "assessment to a risk based, digital approach which will increase the capability to identify fraud" by introducing "risk-based approaches to application assessments, resulting in automation of low risk applications and more time for in depth examination of higher risk cases." Contracts have been awarded to **Kainos Ltd** and **Equal Experts** for the delivery of this project.[58]

In total, we identified 47 companies providing services or specialists to the Home Office since the end of 2016 on its various data-driven transformation plans, via awards listed on Contracts Finder and the Digital Marketplace.

---

[49] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specia ists/opportunities/1227
[50] https://ict.po ice.uk. See also https://assets.pub ishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/512164/po ice-ict-company.pdf
[51] https://ict.po ice.uk/wp-content/uploads/2017/01/Programme-press-version.pdf
[52] https://ted.europa.eu/udl?uri=TED:NOTICE:121962-2018:TEXT:EN:HTML
[53] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/720850/Home_Office_Biometrics_Strategy_-_2018-06-28.pdf
[54] http://bidstats.uk/tenders/2018/W51/693436534
[55] http://bidstats.uk/tenders/2018/W48/692060517
[56] http://bidstats.uk/tenders/2018/W48/691863928
[57] http://bidstats.uk/tenders/2018/W11/674955440
[58] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specia ists/opportunities/2400; https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specia ists/opportunities/4887; https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specia ists/opportunities/4781.

Many other tenders on the Digital Marketplace are listed as closed but no supplier is given. These are omitted from the list below, although the advertisements still give useful indications of work that the department anticipates doing.

The 47 companies are:

Accenture	Fujitsu Services	Netsource
AHE Partnership	Glue Reply	PA Consluting Services
Amberlight Change	Gofore	People Source Consulting
Atos	Hortonworks	Perago-Wales
BAE Systems Applied Intelligence	IBM	Purely Digital Services
Blue Lights Digital	IdentityE2E	Scott Logic
Capgemini	Interact Consulting	Scrumconnect
CGI	Investigo	Softcat
Chaucer Consulting	Kainos Software	Sopra Steria
Computer Task Group	LA International Computer Consultants	SVGC
COMXPS	Leidos	Syrosand Consulting
Deloitte	Marjolo	Teradata
Digi2al	Mastek	Triad Group
Engine Partners	Medley Business Solutions	UKCloud
Equal Experts	Methods Business and Digital Technology	Zaizi
Ernst and Young	Morpho	

Below is a partial list of Home Office awards identified during our research, ordered by category (DSAB and associated, NLEDP, HOB, HMPO) and start date.

We also include a breakdown of the Home Office's G-Cloud purchases by supplier.[59] As mentioned earlier, government transparency data generally doesn't reveal which services were bought.

[59] Because our searches are based on keywords and because these keywords are not app ied consistently by the buyer, where each contract sits within the Home Office's own structure is not always visible. For example, we do not try to unpick whether all "visa" or all "Borders & Immigration Technology Programme" contracts fall within DSAB or are separate.

Digital Services at the Border / Immigration Technology Portfolio

Dec. 2016

Digital Services at the Border require a CESG Tailored Assurance Service (CTAS) Evaluation. CGI, £55k[60]

A Border Force officer is required to perform detailed checks for incoming and outgoing individuals at the border. “Digital Services at the Border (DSAB) requires a inter domain gateway for validation of information across a security boundary. … The Home Office has requirement that the inter domain gateway undergo a CESG Tailored Assurance Service (CTAS) Evaluation. … Without this assurance the new DSAB platform will not Go Live."

DSAB Cloud hosting. UKCloud, £810k[61]

DSAB - AFTC - AFTC IaaS Hosting. UKCloud, £810k[62]

Jan. 2017

DSAB Technical Architect/Developer (AFTC-1). Sopra Steria, £336k[63]
"Work on the Digital Services at the Border programme (DSAB) supporting the AFTC project to work with external 3rd parties to agree and design the approach for receiving and integrating their data within the DSAB data store.” Must support cloud hosting and understand impact of cloud on design.

DSAB Java Developer (AFTC-2). Sopra Steria, £336k[64]
"Work on the Digital Services at the Border programme (DSAB) supporting a data ingestion and management project. This will involve writing code to transform and manipulate data across formats as well as its potential impact on the Hadoop big data platform. There is also a requirement to work with an old project in order to understand requirements and existing behaviour.” Requires expertise in Java 8, Camel, Swagger, SpringBoot, SpringData, Docker, JUnit, Cucumber, micro services architecture, REST, GIT, PHP, Zend, MySQL. “Must be able to retrieve data from micro-services architecture, manipulate data between different forms and understand how this affects the data storage and retrieval."

DSAB - Angular/PHP developer (AFTC-3). Sopra Steria, £312k[65]

[60] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specia ists/opportunities/1414
[61] https://www.contractsfinder.service.gov.uk/Notice/48049cff-ab1b-44ea-b87a-3abd28dd12cf
[62] https://www.contractsfinder.service.gov.uk/Notice/afdf5c6e-75b2-4660-a85d-06875de810c3
[63] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specia ists/opportunities/1203
[64] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specia ists/opportunities/1204
[65] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specia ists/opportunities/1209

"Developer to work on a User Interface project as part of the DSAB Programme. This will involve writing code in the new UI project, as well as working with the old UI project to understand requirements and behaviours."

DSAB - BX - F5 specialist hardware (plus 1 yr support). Softcat, £149k[66]

DSAB Technical Architect / Developer - AFTC1. Sopra Steria, £336k[67]

DSAB PHPDeveloper - AFTC3. Sopra Steria, £312k[68]

DSAB - Client Side Project Review. Scott Logic, £1[69]

DSAB-DOS-Specialist1-AFTC3. Sopra Steria, £393k[70]

DSAB-DOS-Specialist1-AFTC1. Sopra Steria, £424k[71]

Feb. 2017

DSAB - AFTC - Technical Lead Developer. Sopra Steria, £336k[72]

DSAB - AFTC - PHP Developer. Sopra Steria, 312k[73]

Mar. 2017

Border Force (BSTP) Project Management Service. Methods Business and Digital Technology, £1.7m[74]

"Supporting the programme team managing integration of new digital platforms into the current systems whilst managing changes into existing legacy systems in parallel. … The Home Office and Border Force UK are transforming services with the introduction of a number of new digital systems. Migration to these new systems is complex demanding a series of transitional hybrid states from old to new. During this transition period the legacy systems are maintained and developed using key technologies including C++, C# .NET, Java and SQL server, IBMs large infrastructure products such as Websphere Message Broker, Business Objects and Oracle. … The existing team is a blend of internal HO staff supplemented by external consultants. The legacy systems are maintained by multiple external Systems Integrators whilst the new digital platforms are typically developed by in-house teams."

---

[66] https://www.contractsfinder.service.gov.uk/Notice/ba9a5cdf-fe69-4049-b636-2c26f56fbe5a
[67] https://www.contractsfinder.service.gov.uk/Notice/e82b0dce-ca98-4e68-975b-0c6062cafc97
[68] https://www.contractsfinder.service.gov.uk/Notice/e19c2bbc-637d-4ccf-914e-39b963325704
[69] https://www.contractsfinder.service.gov.uk/Notice/8f2622e6-98ca-4899-b839-2ac192100855
[70] https://www.contractsfinder.service.gov.uk/Notice/0609766c-9886-45cd-9bc5-2b94282a292b
[71] https://www.contractsfinder.service.gov.uk/Notice/50413e9a-3266-4505-89cc-9b0590cec367
[72] https://www.contractsfinder.service.gov.uk/Notice/3435166c-8381-48fb-8336-468e4a6d7ec5
[73] https://www.contractsfinder.service.gov.uk/Notice/e5ee9911-ae37-4e93-bb29-125a9263e805
[74] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specia ists/opportunities/2174

Apr. 2017

DSAB - BX - x1 e-Learning Content Developer. People Source Consulting, £150k[75]

DSAB - E-Learning Content Developer. People Source Consulting, £75k[76]

DSAB - SL Variation (IR35 remediation). Scott Logic, £250k[77]

May 2017

Borders Business Architect (BBA1). Marjolo, £84k[78]

"Experienced Business Architects working across the Border Force business domain to design target operating models and associated business architecture artefacts in order to realise the business transformation by 2020."

June 2017

DSAB AFTC Project Test Engineering. IdentityE2E, £1.2 to 2.5m[79]

"Quality Assurance and Testing (QAT) requirements and arrangements needed for the provision of test engineering services to support the delivery of the Advanced Freight Targeting Capability (AFTC) & Advanced Border Control (ABC) projects"

July 2017

DSAB Delivery Manager Roles x3. Marjolo, £660k[80]

Aug. 2017

Borders Java Developer. Triad Group, £286k[81]

"Contribute to the development of the Borders AFTC service working in an Agile delivery environment. The role works on a data ingestion/management project, involving writing code to transform and manipulate data across formats, into storage with potential impact on the HBASE big data platform, providing data to consumers via micro-services." Requires expertise in Java 8, Spring / Spring Boot and spring data rest, REST, TDD, BDD (specifically

[75] https://www.contractsfinder.service.gov.uk/Notice/339163cf-0be5-406e-a778-5ad695a310f8
[76] https://www.contractsfinder.service.gov.uk/Notice/09f511d3-6e9e-4ad0-bc9f-344141c0902d
[77] https://www.contractsfinder.service.gov.uk/Notice/6f4cfc73-7af8-404d-8815-a0616c4fb052
[78] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specia ists/opportunities/2180
[79] https://www.contractsfinder.service.gov.uk/Notice/c63c0248-16bf-4402-b720-9474f0c8e999
[80] https://www.contractsfinder.service.gov.uk/Notice/63c60485-dc0a-4bf7-a5b6-9382628654a2
[81] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specia ists/opportunities/2615

Cucumber), Apache Camel, Microservices architecture, GitHub / GitLab, use of branching and merging strategies, Clean code.

Borders Big Match Developer. Triad Group, £286k[82]

"Contribute to the development of the Borders AFTC service working in an Agile delivery environment. Working closely with data on-boarding, data analytics and devops teams within the project. Role focuses on Infosphere 'Big Match for Hadoop'." Requires expertise in Infosphere Big Match for Hadoop, Hadoop, HBase, MapReduce, Java 8, REST, TDD, BDD (specifically Cucumber), Microservices architecture, GitHub / GitLab, use of branching and merging strategies, Clean code.

DAC-030 - C8367 NGDA Product Manager. Syrosand Consulting, £376k[83]

"The project is providing analytic processing on data arriving within a window of opportunity in order to support Border Force business needs. Specifically, to analyse the data for subjects (both people and goods) approaching the UK border and provide staff with analytics driven insights to allow intervention where necessary. ... Delivery of a number of operational trials that if successful, and provides the required level of benefits will proceed to full implementation using the analysis of data for goods and people, which is on the whole, currently a manual process. The delivery of a data analytics capability to Border Force will enhance their ability to identify targets that produce a positive result, in terms of detection and intervention."

DSAB Java Developer. Triad Consulting & Solutions, £286k[84]

Nov. 2017

DSAB 2nd and 3rd Line Support. Leidos, £5m[85]

G-Cloud 9 DSAB Data Platform - Hadoop Services Support. New Contract. Teradata UK, £1.8m[86]

Case Working Delivery Partner 2. Accenture (UK), £9.5m[87]

"The goal is to replace legacy systems by building a new extensible and reusable framework to meet the needs of immigration and asylum claims within UK Visas and Immigration (UKVI). … As a result of delivering the framework a number of business critical legacy systems will be replaced. … The Portfolio is building services to transform the way the Home Office (HO) manages immigration into the United Kingdom. The key outcome is to build a core Caseworking system for use by Home Office staff, and replace the existing ASYS, BRP,

[82] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specia ists/opportunities/2616
[83] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specia ists/opportunities/4908
[84] https://www.contractsfinder.service.gov.uk/Notice/1451984b-3f85-4557-9660-0283a3b23482
[85] https://www.contractsfinder.service.gov.uk/Notice/de8c1782-8b6b-4c48-848c-716bf871acf7
[86] https://www.contractsfinder.service.gov.uk/Notice/2982c21e-fb33-4281-85c0-6474d396ac61
[87] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specia ists/opportunities/5771

CID, ICW and PBS systems. These systems collectively support several thousand concurrent users, and mange decision making on hundreds of thousands of Immigration and Visa cases per year. Key deliverables include: developing configurable technology modules tailored to a specific business service need; automating existing manual, paper-based processes; and supporting decision-making to drive consistency."

Users need to "Resolve the identities of individuals applying for Visa and Immigration products; Manage information about individuals and their cases; Perform and record the outcomes of checks, including calculations (eg. to assess whether a threshold in the Points Based System is met); Record decisions made about cases and the fulfilment of those decisions (eg. to provide a Visa); or provide asylum support [to] ensure applications are completed within agreed SLAs."

Dec. 2017

Client Side Partner for Immigration Technology. PA Consulting Services (UK), £10m[88]

"Immigration Technology has a requirement for a partner to provide large scale IT programme and project delivery expertise and guidance. … We are helping to deliver the future of immigration. Users include the public applying for entry to or the right to remain in the UK and police, security and enforcement teams controlling immigration and securing UK borders and detecting criminal behaviour within the Home Office and wider Government Agencies."

Business Change Service for UK Visas and Immigration & Immigration Enforcement. Atos IT Services UK, £4.9m[89]

"Business-Change Service provision for an existing-portfolio of in-flight projects, ensuring business-readiness and benefit-realisation."

"UK Visas and Immigration is responsible for making millions of decisions every year about who has the right to visit or stay in the country, with a firm emphasis on national security and a culture of customer satisfaction for people who come here legally. Immigration Enforcement exists to reduce the size of the illegal population and the harm done to the UK. … UKVI users include the public applying for entry to or the right to remain in the UK. IE external users comprise the illegal migrant population. UKVI and IE have internal users of approximately 7,000 and 5,000 respectively."

Digital Customer Journey - Access UK. Deloitte, £23.5m[90]

"The Immigration IT Portfolio is delivering around 70 projects which impact over 20,000 Home Office staff and millions of customers worldwide. We are looking for an AGILE partner to deliver a range of internal and public facing digital immigration services from discovery to

[88] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specia ists/opportunities/4753
[89] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specia ists/opportunities/5088
[90] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specia ists/opportunities/5428

live. … The Home Office is currently delivering digital immigration services to customers with a mixed delivery team. The medium-term plan is to build an in-house AGILE 'Digital Hub' development capability, located in Sheffield. Until this internal capability is established, the Home Office has a need to access AGILE development teams to continue building, testing and iterating customer-facing services, and then transition these skills and knowledge to the internal capability."

IPT L2/L3 Live service support /change. Mastek (UK), £9.5m[91]

"The Immigration IT Portfolio is delivering around 70 projects which impact over 20,000 Home Office staff and millions of customers worldwide; building a range of internal and public facing digital immigration services to our customers. We require a partner to provide interim Level 3 application support for these digital services. Our medium-term plan is to create and build an in-house capability, we are therefore looking for a partner who can transition the service and knowledge to this capability."

Jan. 2018

Borders and Immigration Advanced Border Crossing Delivery Team. BAE Systems Applied Intelligence, £4.9m[92]

"The Digital Services At Border (DSAB) Programme has a requirement for a Delivery Partner to supplement the existing team(s) and supply the capability to deliver discrete work packages aligned to the DSAB programme roadmap, specifically focusing on the delivery of a digital capability to allow Border Force, Police and other key stakeholders to identify and assess threats who are about to travel to, and from, the UK." Initial workpackages: End to end design that aligns to the enterprise level DSAB architecture, including IdAM and Security; Large scale Data Ingest; Information and Analytics.

User-Centred Design. Digi2al, £5m[93]

"Immigration IT Portfolio. The Home Office currently lacks capability in user centred design and the services are failing to meet the Digital by Default Service Standard. The supplier will support in-house delivery teams to build services which meet the Digital Service Standard, reduce duplication of effort, reduce failure demand and transfer knowledge and skills to Civil Servants."

Digital Passenger Services: Developer Outputs. Engine Partners UK, £2.3m[94]

"The initial requirement is to contract for provision of developer services for the Digital Passenger Services Programme, but potentially wider Border Force IT Portfolio."

---

[91] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specia ists/opportunities/5478
[92] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specia ists/opportunities/5403
[93] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specia ists/opportunities/5224
[94] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specia ists/opportunities/5439

Feb. 2018

Digital Services at the Border - DevOps. Capgemini UK, £2.3m[95]

"Increase the DevOps capability".

Mar. 2018

Case Working Delivery Partner 1. PA Consulting Services (UK), £9.5m[96]

(See description for Accenture, Case Working Delivery Partner 2, Nov. 2017.)

Border Crossing Solution Architect. Capgemini UK, £351k[97]

"Supporting the Border Crossing project which is delivering new border control software to the UK borders and a replacement for the current watchlisting systems."

Apr. 2018

Digital Services at the Border Programme - Development Capability. Capgemini UK, £4.3m[98]

Digital Services at the Border Programme - Business Analysis Capability. Capgemini UK, £3m[99]

May 2018

Digitisation of Processes for Border Force Officers. Zaizi, £4m[100]

"Border Force Officers use a suite of legacy tooling, this work will digitise process around most common use cases – typically of the nature of data capture, workflow and present a Management Information dashboard. … This work is required to address many of the deficiencies in day to day operational use and digitise process alongside the rollout of new technology – new laptops, smartphones and integration with cloud based solutions. … Much of the incumbent solutions are locally built solutions based on Excel and Access which do not scale to meet the present multi-user requirements or meet our ambition for becoming a more data driven organisation."

[95] https://www.contractsfinder.service.gov.uk/Notice/5caa2d89-98d8-4bc3-8bc6-c4284762130c
[96] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specia ists/opportunities/5768
[97] https://www.contractsfinder.service.gov.uk/Notice/13c2219a-7033-441b-9efc-0cb5fb73d8b9
[98] https://www.contractsfinder.service.gov.uk/Notice/079a9949-8fe1-41a4-99d3-d9f98ad4c7e4
[99] https://www.contractsfinder.service.gov.uk/Notice/8a179592-671e-4ac8-9319-8caf1d6997ca
[100] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/6309

DSAB Cloud Support Professional Services. Hortonworks UK, £1.7m[101]

"Provision of specialist Hadoop and Big Data Specialist Team"

DSAB - Scrum Master 3. Scrumconnect, £195k[102]

July 2018

DSAB - Scrum Master 1. Triad Consulting & Solutions, £216k[103]

Aug. 2018

Border Systems Technology Projects (BSTP) - Project and Programme Management. Capgemini UK, £3m[104]

"BSTP delivers programmes and projects that support security at the border. Existing projects include upgrades to reporting mechanisms, new software functionality and consolidation work, but the team can be flexed to support any new requirement. The supplier will provide project management capability, release management and planning services to ensure successful delivery of the projects."

Oct. 2018

Border Systems Technology Project (BSTP) - Critical Systems Project Delivery Capability. Methods Business and Digital Technology, £4m[105]

"Borders IT require a specialist Delivery Partner for critical systems project delivery on the Semaphore and Warnings Index programmes, covering programme and project management, architecture and data modelling. Both are critical national systems protecting the border - more detail will be provided at the next stage."

[101] https://www.contractsfinder.service.gov.uk/Notice/9db1a902-fefc-426d-9378-e198bfeb4296
[102] https://www.contractsfinder.service.gov.uk/Notice/448f1d90-f6ab-492e-9c7e-27f6447134b9
[103] https://www.contractsfinder.service.gov.uk/Notice/d29bce97-4044-4f2b-a236-d67abf1a6e43
[104] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/7573
[105] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/7539

National Law Enforcement Data Programme

Sep. 2016

Delivery Manager - Data Migration Project Manager. Chaucer Consulting, £165k[106]

"Requires a deep understanding of all aspects of migration from a Natural/Adabas database mainframe environment. The specialist will oversee the data migration strategy, and document relevant business logic and rules in order to migrate to a new system. The candidate will be responsible for ensuring a smooth transition to the target platform managing trial runs, data reconciliation, data mapping, readiness assessments."

Oct. 2016

Business Analyst. Medley Business Solutions, £176k[107]

"Engage with the law enforcement community, other Home Office Digital, Data and Technology (HODDaT) programmes, and other programme streams, to progress the design of the integrated services across LEDS."

Technical Transition Architect. Medley Business Solutions, £176k[108]

"Technical transition architect, responsible for defining the transition solution from a complex Adabas/Natural mainframe environment to a new platform."

Nov. 2016

Planning Delivery Manager. AHE Partnership, £157k[109]

"Plan, structure and lead the planning of the projects ... Manage the interface between projects highlighting both the dependencies and the interfaces between stakeholders."

Jan. 2017

Stakeholder Engagement Manager. Triad Group, £127k[110]

"Work collaboratively across National Law Enforcement Data Programme (NLEDP) to deliver the programme vision, developing and maintaining a stakeholder engagement strategy and plan to ensure consistent and coherent engagement across NPDP's diverse stakeholder landscape."

---

[106] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/591
[107] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/868
[108] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/874
[109] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/948
[110] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/1175

PND Service Architect. Methods Business and Digital Technology, £151k[111]

"Service Disaggregation of an existing contract. To work with the existing team to define, document and agree with all stakeholders the approach, implementation and ongoing service architecture. You will own the service disaggregation including security considerations and help shape the services, metrics and operating model for the application.

Feb. 2017

Application Development Service. IBM UK, £12m[112]

Building and supporting data services to support law enforcement and other agencies; highly automated application delivery pipelines; access portals; API gateways supporting fixed and mobile channels.

"The Home Office has existing data systems whose contracts begin to expire in 2019. These systems are expensive to operate and maintain, and difficult to change to accommodate the rapidly evolving law enforcement landscape. Additionally, some systems and components are approaching end-of-life, or end-of-support. This procurement will enable NLEDP to replace these systems with the Law Enforcement Data Service (LEDS). LEDS will ensure continuity of business services, and act as a platform for innovation to transform the way the HO manages and supplies data services to Law Enforcement Communities and other authorised Agencies throughout the UK and internationally. … Work packages will focus upon access portals, web applications, API gateways, Master Data Management, Entity Recognition, messaging, micro-services, security and auditing, container technologies and SQL and NoSQL databases. … Users include 43 police forces, Border Control / Immigration, government departments and agencies, Disclosure Services, Criminal Justice System, Offender Management Services. User requirements include the provision of capability to: check an individual's identity, offending history, status, and location; analyse data to identify links between people, objects, locations and events; set up automated alerts for new or changed data and events of interest."

User Researcher. Amberlight Change, £146k[113]

"Engage directly with Service Designers and Business Architects, operational police officers and the public in order to create local, regional and national insight in the design of service, process and organisational structure."

Relationship Manager. Methods Business and Digital Technology, £143k[114]

"You will play an integral role in securing access to data sources necessary to fulfil the wider data sharing ambitions of the programme and will own relationships/engage with data

[111] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/1140
[112] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/1227
[113] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/1218
[114] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/1173

owners ensuring data sharing requirements/implications are understood whilst ensuring plans are put in place to address these in line with Ministerial direction."

June 2017

Non Standard Service (NSS) Delivery Project Manager. Purely Digital Services, £151k[115]

"Identify and manage relationships and inter-dependencies between projects within a complex commercial environment of multiple suppliers, 3rd parties and partners."

July 2017

Solution Designer. LA International Computer Consultants, £249k[116]

Transition Service Packages, Transition Risk Mitigation and Rollback, Interface and Batch treatment plans, Data Replication, Reconciliation.

Service / Business Architect. Perago-Wales, £324k[117]

"Shaping how the future of LEDS will operate in a BAU environment."

Product Manager. Computer Task Group (UK), £581k[118]

"The programme is comprised of a mixed team of Civil Servants and client side consultants supplemented with SMEs both client and supply side. The programme team also comprises seconded operational and retired Police Officers. The programme also utilises one large supplier (Application Development Support Partner) for most of the build activity associated with LEDS and a collection of small SMEs focused on activities such as data migration, testing and data quality. … One of NLEDP's key objectives is to determine the future need for new services to make national data easy to access by frontline officers to inform local policing decisions. As Product Manager, you will support NLEDP delivery by analysing and defining requirements to and establishing core LEDS products that work together to provide the overarching service.”

Relationship Manager. Computer Task Group (UK), £581k[119]

"As Relationship Manager, you will support the NLEDP leadership team in securing access to data sources necessary to fulfil the wider data sharing ambitions of the programme. You will engage with data owners to ensure that data sharing requirements and their implications are understood and that plans are put in place to address these in line with Ministerial appetite and direction."

---

[115] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/2082
[116] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/2374
[117] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/2480
[118] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/4563
[119] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/4565

Business Analyst. Computer Task Group (UK), £490k[120]

"Support the NLEDP delivery teams by analysing and defining requirements to make sure that the projects that deliver them are built and run to align to the service vision, meeting business and user needs."

Project Manager. Computer Task Group (UK), £581k[121]

"Support the NLEDP leadership team in delivering outputs necessary to fulfil the wider outcome delivery objectives of the programme".

Service Transition & Programme Support. Ernst & Young, 4.5m[122]

"There is a need to manage the transition of services being developed to a BAU state. To do this, the programme requires a Service transition team, comprising of a Service Transition Lead and at least 2 Service transition Managers to deliver outcomes such as Service Design Packages, Impact assessments, Service Transition governance."
(See also Feb. 2017, IBM award.)

Aug. 2017

Governance Lead. SVGC, £259k[123]

"You will scope business level governance aspects of Sustainment including establishing a comprehensive map of current and to-be governance and identify governance and processes required to agree these provisions with the business."

Service Manager. Netsource, £222k[124]

"You will support the Sustainment Lead by establishing a clear plan for an ITiL aligned service organisation. You will outline scope and requirements for delivering the proposed Service Architecture. You will support planning for Sustainment and act as an interim service manager for any services not yet transitioned to operations."

Oct. 2017

Project Manager - PNC. Chaucer Consulting, £374k[125]

"Support the NLEDP leadership team in delivering outputs necessary to fulfil the wider outcome delivery objectives of the programme."

---

[120] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/4568
[121] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/4570
[122] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/4567
[123] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/2508
[124] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/2509
[125] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/5090

Jan. 2018

Interaction Designer for Police Intelligence. Triad Group, £275k[126]

"The interaction designer will be working to refine a prototype based on feedback from users who work as intelligence analysts in police forces. The interaction designer will then work on scaling the product for live and may assist on other LEDS products. … We are specifically exploring how to replace the service that intelligence staff in all UK Police Forces use to find intelligence at a national level with a new UI supported by modernised search capabilities."

Apr. 2018

Interaction Designer for Police Intelligence. Interact Consulting, £158k[127]

"Lead the design of a national police intelligence service based on police officers' user needs. The product will be used by Police Forces on a daily basis to manage national intelligence and is hoped to set a design pattern for other search services. … LEDS will replace the service that police users find intelligence on at a national level with a new UI supported by modernised search capabilities."

Service Transition Manager-1. COMXPS, £283k[128]

"Mobilisation and management of resources, facilitating and tracking progress across service suppliers (internal HO teams and external suppliers)."

Service Architect. Computer Task Group (UK), £157k[129]

"Engage with key stakeholders to develop service model, service architecture and support service transition plans for the multiple operating states that NLEDP will deliver."

Systems Integration Team. Glue Reply, £4.5m[130]

"The transition architecture and approach seeks to minimise mandated external change whilst sustaining existing legacy interactions. Each legacy interface requires an interface approach/treatment/definition and dependency roadmap that is agreed with the respective integration partner or partners. Partners typically are other government departments, law enforcement agencies, and commercial vendors of services and products. … There is an incumbent supplier providing the current team that will transition out. There will be a handover phase during the first 1-2 months of the contract to the new supplier. … LEDS must support its UI and API interactions by using information flows to external systems. It

[126] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/5576
[127] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/6158
[128] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/6151
[129] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/6152
[130] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/5999

shall also provide system interfaces to enable approved external organisations to initiate interactions with LEDS data. This integration uses several Enterprise Integration Patterns. The target for NLEDP is to reduce the numerous legacy integration points but, to avoid significant software and business change, NLEDP aims to support the existing information flows until external partners can modernise their interfaces alongside the necessary business change."

May 2018

Service Transition Manager-3. BAE Systems Applied Intelligence, £232k[131]

"Manage Service Transition Plan. Mobilisation and management of resources, facilitating and tracking progress across service suppliers (internal HO teams and external suppliers)."

Migration Solutions Architect. People Source Consulting, £345k[132]

"Specialist will lead on the discovery of the data migration requirements, development of data migration strategies, and design of technical solutions for migrating data from current systems into their replacements, including both business applications data and end-user data."

July 2018[133]

Data Migration Project Manager. Chaucer Consulting, £850/day[134]

“The specialist will ensure a smooth transition from source to target platform, managing trial runs, data reconciliation, data mapping, readiness assessments and informing the development of replacement services. Significant experience required in Data Architecture, data quality, Natural/Adabas and robust Project Management, able to manage business and technical aspects of migration. … The role requires a deep understanding of migration from a Natural/Adabas database mainframe environment. The specialist will oversee the data migration strategy, and document relevant business logic and rules in order to migrate to a new system.”

Oct. 2018

Learning Partner Delivery. Blue Lights Digital, £2m[135]

"The Law Enforcement Data Service (LEDS) will replace two critical national police IT services provided by the Police National Computer (PNC) and Police National Database

[131] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/6592
[132] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/5467
[133] Seemingly mistakenly isted in DMP as July 2017.
[134] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/6685
[135] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/6853

(PND). The Programme is seeking to appoint a learning partner to design, manage and deliver training and learning materials to the User community."

Technical Transition Architect. Medley Business Solutions, £410k[136]

"Requires a deep understanding of system-to-system interfaces with experience of transition and migration from a mainframe architecture."

Dec. 2018

LEDS Interaction Designer. Gofore UK, £302k[137]

"The interaction designer will work to create a new national police intelligence and data service based on user needs. The product will be used by police officers and operational staff in UK Police Forces to manage national intelligence data and provide information pertinent to the investigation of crime."

[136] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/7501
[137] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/7320

Home Office Biometrics (HOB)

Feb. 2018

Home Office Biometric Matcher Platform and associated Services - Lot 1. Fujitsu Services, £28m[138]

"The Contracting Authority intends to identify an Economic Operator(s) to provide the Biometric Matcher Platform & associated Services (BMPS). BMPS delivers across multiple biometric modalities and data sets (which shall include, but not limited to, immigration, citizenship, law enforcement) as part of Home Office Biometrics (HOB) programme. There are two distinct parts to the BMPS: (i) A technology platform and Biometric Matcher Service Bus; and (ii) Biometric Matching algorithm(s) that are integrated into the platform.
Three separate lots will be used to procure the BMPS.
Lot 1: Matcher Service Supplier (MSS) will be selected to provide Matcher Service Platform (MSP) that includes a Matcher Service Bus (MSB), Matcher service interface, integration of MES(s) and operation of the BMPS."

Mar. 2018

DevOps - Platform Development and 3rd Line Support. Mastek (UK), £5m[139]

"HOB Programme delivers future biometric IT services to various government departments including Home Office departments and Law Enforcement agencies. Consumers use the biometric services to support various business processes for immigration, law enforcement and citizenship. The services run on the platform provided and managed by the DevOps Tooling Platform service. … A number of services are already live including the Biometric Services Gateway and several services are under development including international data sharing capabilities and additional APIs. The DevOps Tooling Platform service will need to take on and manage the platform in support of the existing live services and also transition the work in progress for the in-flight projects. The development tooling and pipeline is already in place in support of the existing IaaS supplier. It is anticipated that the development pipeline and tooling would need to be changed in order to support the transition to a new
IaaS supplier."

DevOps Platform Development and 3rd Line Support. Mastek (UK), £5m[140]

May 2018

[138] https://www.contractsfinder.service.gov.uk/Notice/21319a82-4a57-4b62-8eff-57a320a2328e
[139] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/5673
[140] https://www.contractsfinder.service.gov.uk/Notice/e0786905-1130-4ec7-8d73-f87ec9c877ca

Home Office Biometric Matcher Platform and associated Services - Lot 2. Morpho UK, £7m[141]

(See Matcher Platform, Feb. 2018.)
"Lot 2: Matcher Engine Software (MES) (minutiae based) Supplier will be selected to provide matching of Tenprint, Palm and Latent marks including support for verification, search and de-duplication of biometric records."

Aug. 2018

Biometrics Managed Service Provider (MSP). Investigo, £4.5m[142]

“The Managed Service Provider (MSP) will work with the Buyer's Programme team to deliver a large, complex programme converging IT systems into a cohesive, cost effective, user centred service. This will include elements of legacy outsourced Services (typically biometric systems) being transition into new and mixed delivery architectures including but not limited to Cloud-based services, Buyer managed environments, and/or 3rd party technical infrastructures and/or capabilities.”

Jan. 2019

Biometric Technical Services. Identity E2E, £4.9m[143]

[141] https://www.contractsfinder.service.gov.uk/Notice/87f8f797-acb9-43d1-8600-f7f8fa04cc9f
[142] https://www.contractsfinder.service.gov.uk/Notice/549867be-6441-4f6b-b5f0-413e09d43141
[143] https://www.contractsfinder.service.gov.uk/Notice/ad31dfb0-1037-4c8b-a2e2-18c0aae79f01

HMPO Digital Application Processing

July 2017

Delivery partner for online passport application services. Kainos Software, £3.2m[144]

"HM Passport Office has an organisational goal to create an integrated, digital-by-default service for customers who need to apply, renew or replace passports. This will reduce the 35 million documents and pieces of paper that HM Passport Office and its partners handle annually to the single digit millions by 2020, by receiving over 90 per cent of passport applications online. … The existing service accommodates a single customer group - adult renewals aged 26 years and older. HM Passport Office's plans are to extend this to all customer groups, both in the UK and abroad, in order to achieve the strategic organisational goal of 95% transactions online by 2020. The current delivery approach needs to be scaled to support this ambition, without sacrificing the user-centered approach already established."

Oct. 2017

Product & Delivery Team. Equal Experts UK, £4.5m[145]

"HMPO's 2020 vision moves passport applications online for the majority of customers and shifts assessment to a risk based, digital approach which will increase the capability to identify fraud. The DAP project is responsible for building the capability to deliver an in house case working, workflow and checking capability to support the production of passport applications, currently the responsibility of an external provider. With HMPO's contracts with existing strategic partners coming to an end in 2019, the DAP service has to be ready to take over all application processing by this date. … To achieve the HMPO Transformation agenda, DAP needs to introduce risk-based approaches to application assessments, resulting in automation of low risk applications and more time for in depth examination of higher risk cases."

Nov. 2017

Delivery Team Supporting Casework Team. Equal Experts UK, £2m[146]

See above, Oct. 2017. "As a user the HMPO passport staff need a system that automates passport applications where safe and practical to do so. To enable them to concentrate on the more complex cases that can never be automated."

[144] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/2400
[145] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/4887
[146] https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/4781

G-Cloud Purchases

The Home Office is the largest spending government department on G-Cloud, with £447m of purchases disclosed by the government from January 2016 to December 2018. (This is almost double the G-Cloud spend of the next-highest-spending department, HMRC.) Over that period, the Home Office has bought services from roughly 180 firms, listed below in descending order of cost.[147]

Capgemini Uk Plc	£48,329,405
BAE Systems Applied Intelligence	£31,982,502
Mastek UK	£29,820,050
Deloitte LLP	£29,695,521
IBM United Kingdom	£22,686,931
PA CONSULTING SERVICES	£22,584,579
Amazon Web Services Inc	£22,507,224
Accenture (UK)	£19,949,054
FDM Group	£12,377,301
6point6	£12,131,810
SUNGARD AVAILABILITY SYSTEMS (UK)	£10,846,132
Methods Professional Services	£10,431,666
PA CONSULTING SERVICES LIMITED	£8,957,822
UKCloud	£8,955,950
Engine Partners UK LLP	£8,352,191
Redweb	£8,220,727
Leonardo MW	£5,751,940
Certus Solutions	£5,232,876
Amazon Web Services EMEA SARL	£5,055,866
Kainos Software	£4,235,652
NETbuilder	£4,215,393
Roke Manor Research Limited	£3,984,683
Alpine Resourcing Limited	£3,588,987
Big Data Partnership	£3,577,210
QA Consulting Services Limited	£3,574,665

[147] Data presented in its original form. Some firms are listed under multiple names, but we have not tried to deduplicate these, nor have we adjusted capitalisation for greater consistency. Note that companies are often not listed by the same name in the G-Cloud transparency data as they are on the G-Cloud search site.

Atkins	£3,545,780
Roke Manor Research	£3,389,067
Fujitsu Services	£3,280,719
DataLynx	£3,220,039
DataLynx Limited	£3,072,517
Alpine Resourcing	£3,011,467
PricewaterhouseCoopers LLP	£2,987,805
NETCOMPANY UK	£2,948,388
QA Consulting Services	£2,909,437
ITM Communications	£2,755,995
Scott Logic Limited	£2,751,320
Fincore	£2,423,862
LA International Computer Consultants	£2,314,289
Be Informed BV	£2,119,607
ITM Communications Limited	£2,007,523
SFW	£2,000,532
Chaucer Group Limited	£1,942,523
NETCOMPANY UK LIMITED	£1,915,382
PARITY RESOURCES	£1,666,508
The Strategy & Architecture Group	£1,648,152
WorldReach Software Corporation	£1,532,571
Think Big, A Teradata Company	£1,500,440
Veran Performance	£1,485,146
Ernst & Young LLP	£1,476,432
Fincore Limited	£1,469,560
Actica Consulting	£1,403,435
Identity E2E	£1,375,909
Invotra	£1,332,210
MEDLEY BUSINESS SOLUTIONS	£1,209,425
L-3 Communications ASA LImited	£1,200,322
Mozaic-Services	£1,192,318
Mindful Contract Solutions Limited	£1,164,639
Hive IT	£1,159,849
Triad Group Plc	£1,126,995
L-3 Communications ASA	£1,097,716

Scott Logic	£1,076,068
Civica UK	£1,017,403
Westhill Commercial	£1,007,991
Equal Experts	£939,950
TRUSTIS	£864,190
Dome Recruitment	£818,211
IPL Information Processing	£798,626
Sopra Steria	£788,463
Trustmarque Solutions	£776,935
Microsoft UK	£714,117
Zaizi	£661,401
BRAMBLE HUB	£656,766
Methods Business and Digital Technology Limited	£637,734
SPECIALIST COMPUTER CENTRES PLC	£609,371
Westhill Commercial Limited	£577,106
Zeefix Consulting Limited	£565,330
Airwave Solutions Limited	£537,990
FCO SERVICES	£529,215
Trustmarque Solutions Limited	£508,634
Vodafone	£506,272
NCC GROUP PLC	£496,822
NINIAN SOLUTIONS T/A HUDDLE	£469,275
APPVIA	£423,383
Kinegistic	£410,250
Media Measurement Limited	£392,607
Chorus Intelligence Limited	£391,416
Telefónica UK	£359,650
WTG Technologies	£349,076
Border-Systems.com	£334,704
Methods Professional Services Limited	£317,180
F-SECURE DIGITAL ASSURANCE.	£306,106
IO1	£305,140
People Source Consulting	£303,475
Modis International	£301,711
RAINMAKER SOLUTIONS	£301,250

BSI CYBERSECURITY AND INFORMATION RESILIENCE (UK)	£292,706
Clearvision CM 2005	£290,945
scrumconnect	£276,340
Basware Holdings Limited	£267,974
Eduserv	£263,927
DEMYSTIFY PROFESSIONAL SERVICES	£257,505
Altius Consulting	£249,785
Dun and Bradstreet	£224,000
Planixs GRP	£213,000
Sopra Steria Limited	£211,931
KPMG LLP	£206,934
Ninth Wave	£204,228
Hive IT Limited	£193,404
Toplevel Computing	£191,145
Zennor Consulting	£188,446
Computer Task Group UK	£185,384
Allen Lane	£179,335
Assured Information Security	£173,604
Kinegistic Limited	£150,400
BUSINESSWEB SOFTWARE T/A FIRMSTEP	£150,000
Pluralsight, LLC	£145,905
Health HR UK Limited	£138,000
Blackthorn GRC Limited	£137,316
Amazon Web Services UK	£134,539
Nudge Digital	£121,868
Blackthorn GRC	£114,481
BravoSolution	£110,986
Veracity OSI UK Limited	£109,570
OD Consultancy	£106,615
Kimcell	£102,000
Flabba Limited	£101,332
Big Data Partnership Limited	£96,760
Amazon Web Services UK Limited	£95,169
Atamis	£90,917
Talent International (UK)	£90,605

Mindful Contract Solutions	£82,625
Quo Imus	£80,986
NQC	£74,150
Ninth Wave Limited	£70,528
Box Inc.	£69,798
MDS TECHNOLOGIES	£68,650
Certus Advisory	£64,646
Cassidian CyberSecurity	£64,476
UBDS IT CONSULTING	£62,238
REPKNIGHT	£54,000
SALESFORCE.COM EMEA LIMITED	£51,336
CACI	£50,950
Gartner	£50,800
Barrachd	£50,400
McKinsey & Company Inc United Kingdom	£50,000
ENCIRCLE SOLUTIONS	£49,551
MMGRP	£46,654
ARCUS GLOBAL	£44,348
Vldb Solutions	£44,000
Badenoch & Clark	£42,650
ICASEWORK	£42,500
BSI CYBERSECURITY AND INFORMATION RESILIENCE (UK) LIMITED	£42,340
Dyn	£35,978
Unify	£34,200
GB Group plc	£34,000
BEAUMONT COLSON	£31,325
IndigoBlue Consulting	£29,400
MICROSOFT IRELAND OPERATIONS	£27,335
Panlogic	£26,085
INOVEM	£25,069
SmartSurvey	£25,000
EGRESS SOFTWARE TECHNOLOGIES	£24,450
CAPITA BUSINESS SERVICES	£24,133
The Server Labs	£22,998
Twentysix	£19,871

MEMSET	£19,868
PNH CONSULTING SERVICES LIMITED	£18,000
Computer Application Services	£17,875
Mazepoint	£15,400
IB BOOST	£15,042
Basware Holdings	£15,000
Mazepoint Limited	£12,767
MAGIC MILESTONES	£10,500
Advent IM	£10,200
2T Security	£8,100
British Telecommunications Plc	£6,097
Information Risk Management Plc	£5,850
QuoVadis Online Security Limited	£3,435
Digital Accessibility Centre	£3,375
QuoVadis Online Security	£2,580
Q5 Partners LLP	£1,800
Wired Marketing	£1,470
GLOBAL RADIODATA COMMUNICATIONS.	£681

SUBMISSION 7

AI and Public Standards

Rick Turner, Chief Executive Officer, DITTO AI LIMITED
[redacted]
17 May 2019

PUBLIC

t (+44) (0)330 995 0710
e hello@ditto.ai
w ditto.ai

Ditto AI Ltd Jactin House 24 Hood Street Ancoats Urban Village Manchester M4 6WX

Registered office: Leeds Innovation Centre 103 Clarendon Road Leeds LS2 9DF United Kingdom | Company number: 3442563 | VAT number: 767 7004 53

Background

The Committee on Standards in Public Life (CSPL) is conducting a consultation on "AI and Public Standards" and wishes to:

1. Consider whether existing frameworks and regulations are sufficient to ensure that standards are upheld as technologically assisted decision-making is adopted more widely in the public sector,
2. Examine how provisions for standards can be built into the development, commissioning, and deployment of new technologies in the public sector;
3. Consider to what extent the use of artificial intelligence and associated advanced technology has implications for our understanding and formulation of the Seven Principles of Public Life;
4. Make recommendations for how standards can be maintained in the public sector where advanced technologies are increasingly used for service delivery, including best practice guidance and regulatory change where necessary.

The review will focus on those in public roles but also those in the private sector who deliver public services, including front line services such as policing and law enforcement, justice and health.

In this submission, we at Ditto AI wish to draw attention to the features of AI technologies in development today that have intrinsic qualities to serve the 7 principles of public life.

The challenge posed by AI

The seven Nolan Principles – namely selflessness, integrity, objectivity, accountability, openness, honesty, leadership – define the standards of ethical conduct in public life. The use of technology-assisted decision making, especially the use of AI, can make it harder to adhere demonstrably to these principles. This is a timely consultation.

The issue in our view is not so much in technology-assisted decision making as it is in the "black box" nature of most AI that is in use today. The CSPL consultation specifically mentions policing and law enforcement, justice, and health, all of which impact life or death and can therefore not afford errors.

We would like to use a couple of illustrative examples to show how current "black box" AI approaches work, or rather fail to work, in these important contexts.

First, in July 2018, the American Civil Liberties Union conducted a study that compared photos of all federal lawmakers against a database of 25,000 publicly available images. Amazon's facial recognition technology, called 'Rekognition', falsely identified 28 members of US Congress as people arrested for crimes.

Three of the misidentified legislators sent a letter to Amazon's CEO Jeff Bezos stating that there were "serious questions regarding whether Amazon should be selling its technology to law enforcement at this time."

Crucially the AI algorithm could not explain how these matches -- or mismatches -- were made. This lack of scrutability stands in the way of the practice of Nolan principles by public servants, in this instance, law enforcement officials.

As a second example, machine-learning and deep-learning based solutions are in development to speed up diagnosis of a range of ailments from diabetic retinopathy to lung cancer. These are life altering challenges and even a one percent error in diagnosis would be simply unacceptable. Human expertise forms an essential part of the diagnosis process at the moment, but it does not scale and, often being tacit, is not easily transferable for access when the human expert is not available.

What happens when the mismatched persons in the first experiment are not powerful legislators but ordinary citizens who may be treated with less deference and more suspicion? What happens when a person is wrongly diagnosed with cancer and undergoes painful but unnecessary and expensive (in the UK, to the public purse) treatment? Worse, what happens when a cancer diagnosis is missed, and the patient is not treated, and their longevity curtailed?

How are the public servants to demonstrate Objectivity and Accountability in such situations?

Across the world, there is a demand now to regulate AI but there is no clear view on the form and scope of such regulation. As such there is no guarantee that regulation will be the answer to addressing the real and perceived problems with automated decision making and AI.

To serve the Nolan Principles, the AI technologies need to do better than they are currently doing in enabling decision support in the public sector.

Ditto AI's view

We are a British company developing explainable, accountable AI for a responsible future. As such, the Nolan Principles are embedded in our technology.

Specifically, our patented methodology goes beyond big databases, and codifies tacit and unstructured knowledge and expertise, that informs judgment and decisions in complex and nuanced areas of decision making such as policing, law enforcement, justice and health, which are key concerns in this consultation. It can be difficult to translate such deep experience and tacit expertise into clear, objective explanations. This difficulty can in turn hurt the practice of the Nolan Principle of Objectivity.

An AI decision support system using our patented technology provides an explanation for the outputs in plain English as well as an audit trail, effectively show the workings of the algorithm. This "explanation" is inbuilt in our technology and enables both Openness and Accountability, which in turn enable Integrity and Objectivity in decision making.

More crucially, our AI can work with “black box” AI, currently deployed in many organisational processes to aid decision making, to help make it more transparent and accountable.

Our view is that due to its inherent shortcomings, big data led “black box”AI is unsuitable for enabling decisions in the public sector. However explainable AI, which is being developed by us and some other organisations worldwide, addresses the emerging concerns especially those embedded in the Nolan Principles.

Our recommendation to the CSPL therefore is that the public sector should require Accountability and Openness in AI and other technologies being deployed to support automated decision making. Explainable AI is already capable of making Objectivity, Openness and Accountability possible in decision making in complex areas such as policing and law enforcement, justice, and health.

(End of document)

SUBMISSION 8

Committee on Standards in Public Life

Review into Artificial Intelligence and Public Standards

This submission from the Royal College of Physicians (RCP) focuses exclusively on the use and standards of Artificial Intelligence (AI) in healthcare.

Summary

- AI has huge potential to support doctors and enable them to spend more time with patients. The RCP is excited about the future use of AI to support healthcare, improving patient care and the lives of doctors at the same time.

- The use of AI in healthcare is already a reality, and is evolving at a rapid pace. Doctors use AI to assist safe prescribing, calculate risk and interpret investigation results, among other uses. However, many have concerns about the perceived lack of testing of such technology, and the accuracy of the advice it provides.

- In healthcare especially, it is essential to ensure new technology is safe and effective. This can be increasingly difficult when the development of new technology can quickly outstrip the regulatory environment. We should have the same standard of proof for AI as we would for any new technology or drug.

- Regulators, NHS England and NHS Digital should be supported to adapt to a changing environment, develop guidance, principles and appropriate evaluation methods to assess AI, including clinical and patient input where possible and supporting dissemination of their assessment results.

- Many clinicians are unclear where responsibility lies or the legal implications for support given by AI products. Both doctors and patients should be involved with AI development and testing in a meaningful way to address such challenges at the earliest possible stage

Inquiry questions:

1. Whether existing frameworks and regulations are sufficient to ensure that standards are upheld as technologically assisted decision-making is adopted more widely in the public sector

a. Examining the current use of artificial intelligence and associated advanced technologies in the public sector

We asked our members how they use AI technologies in their practice, and representatives from 19 medical specialties responded. 100% of them reported they currently use AI to support safe prescribing, and 68% said they use it to calculate risk or prognosis. Other notable uses of AI are to interpret investigation results (47%), to monitor long term conditions (37%) and to assist diagnosis

(32%). Our members reported the use of AI in these ways improved safety, medicine management and allowed for greater efficiency.

However, our members also have worries about the use of AI technology in healthcare. 84% reported concerns about the perceived lack of testing of new technology, and the accuracy of advice given. Other important issues include whether new technology would ignore patient preferences or interrupt consultations, and if it is based on the latest clinical advice. Similarly, there were concerns over the level of user experience required to ensure the safe use of technology and whether the technology had been developed with the use of high quality datasets.

Doctors are also concerned about the ethics and legal liabilities of using AI and associated technologies.

d. Examining what safeguards and considerations of standards are currently in place in the deployment of AI and advanced technologies within the public sector

The results of our survey show the use of AI is already widespread, and doctors have well-founded questions about new technology. There is a perceived lack of testing of new technology; meaning that although testing may take place, the results are not communicated to doctors.

Support should be given for doctors and patients to engage in the testing of AI and other new technologies in a meaningful way, and the results of such testing should be clearly communicated. This will also encourage up-take, as clinicians and the public can be assured of their quality.

NICE recently produced a detailed Evidence Standards Framework for Digital Health Technologies in healthcare.[7] This sets out evidence standards for the clinical and economic impact of new digital health technologies, and provides insight for innovators to develop the case for the use of their technology in the NHS. This is also useful for commissioners to know what to ask for from developers and to inform commissioning decisions. This is a promising way forward, which the RCP is keen to endorse.

In a similar vein, the Department of Health and Social Care's code of conduct for data-driven health and care technology is another useful resource that was recently released. It sets out 10 principles for ethical development and deployment of AI tools. When used as part of a wider strategy, the code of conduct should contribute to the creation of a trusted environment supporting AI innovations while ensuring safety and ethics remain at the forefront.

The government's vision for the future of digital healthcare describes a plan to introduce a regulatory 'sandbox'. That promises to be able to test, iterate and de-risk promising technologies to support implementation. More detail is needed about how this will be undertaken.

2. How provisions for standards can be built into the development, commissioning, and deployment of new technologies in the public sector

Standards must be built into AI from its earliest stages. AI does not develop in a vacuum, and if precautions are not taken it can reflect and perpetuate societal prejudices. For example, an AI tool to detect skin cancer may be less likely to diagnose patients with dark skin correctly if the data it was trained on comes from mostly fair-skinned patients.

One of the best ways to ensure AI tools in healthcare are accurate and fair is by training them on high-quality data. Most AI tools need to be trained on huge amounts of accurate and reliable data,

[7] https://digitalhealth.london/nice-esf-v2/

and inaccurate and misrepresentative data can lead to poorly performing systems. Ensuring new technologies in the public sector are developed using accurate and reliable data will go a long way to developing rigorous standards.

In healthcare a lack of universal data standards that facilitate data sharing means datasets often end up in inaccessible formats or difficult to access repositories. There must be a focus on creating high quality datasets across the NHS, with everyone who enters data understanding the data they collect is not solely for their use, but will be shared across and beyond their immediate clinical team. On a bigger scale, Sustainability and Transformation Partnerships and Integrated Care Systems should adopt uniform data standards where possible and encourage the development of high quality datasets across patient populations.

Synthetic datasets could be one possible solution to this problem. These are made using a small sample of real patient data to generate a much larger sample of synthetic data which can be used in the development and testing of AI applications. The Medicines and Healthcare products Regulatory Agency (MHRA) has recently secure funding to explore the creation of synthetic datasets. Public Health England has also recently funded the release of a synthetic dataset for the cancer registry.

Both doctors and patients (or their representatives) must be involved in the development of AI and other new technologies to ensure they deliver immediate user benefits and address relevant problems. This means going beyond user-centred design and agile development to embrace co-production of these tools with typical (randomly selected) clinicians and patients, not informatics experts.

4. Make recommendations for how standards can be maintained in the public sector where advanced technologies are increasingly used for service delivery, including best practice guidance and regulatory change where necessary

It is vital that no patient is left behind if they are unable to engage with new digital health technology. Everyone must have equitable access high-quality, comprehensive care, regardless of age, socio-economic status and technological literacy.

AI applications will never replace a fully trained and qualified human doctor, however advanced they become. To maintain standards in the health sector, the development of AI tools must be focused on the routine, automatable tasks that take doctors away from their patients. This will allow doctors to focus on the more human aspects of their job, and spend more time with their patients offering compassionate care, explaining their diagnosis, prognosis and treatment options.

As stated above, any technology used for medical care must undergo rigorous testing, and the same burden of proof should be required for AI tools as regular drugs. To achieve this, more support must be given to regulators so they can continue to adapt to the changing technological landscape of healthcare.

Finally, both doctors and patients must be protected if new technology fails. AI technology is not infallible. Who is responsible when AI makes a mistake, and how can doctors know how much to rely on new AI tools? More discussion is needed around the real-life application of AI and how it is used by clinicians.

About the RCP

The RCP plays a leading role in the delivery of high-quality patient care by setting standards of medical practice and promoting clinical excellence. We provide physicians in the UK and overseas

with education, training and support throughout their careers. As an independent body representing over 36,000 fellows and members worldwide, we advise and work with government, the public, patients and other professions to improve health and healthcare. Our primary interest is in building a health system that delivers high-quality care for patients.

SUBMISSION 9

Why Value Judgements Should Not Be Automated

Rune Nyrup, Jess Whittlestone, Stephen Cave
Leverhulme Centre for the Future of Intelligence, University of Cambridge

Submitted as evidence for the Committee on Standards in Public Life's review into artificial intelligence and its impact on standards across the public sector.

Summary

1. AI technologies are already being used for a number of purposes in public services, including to automate (parts of) decision processes and to make recommendations and predictions in support of human decisions. Increasing application of AI in public services therefore has the potential to impact several of the Seven Principles of Public Life, presenting new challenges for public servants in upholding those values. **We believe AI is particularly likely to impact the principles of Objectivity, Openness, Accountability and Leadership**. Algorithmic bias has the potential to threaten the *objectivity* of public sector decisions, while several forms of opacity in AI systems raise challenges for *openness* in public services; and, in turn, this impacts the ability of public servants to be *accountable* and exercise proper *leadership*.

2. We suggest that public standards need to be supplemented by more **concrete, action-oriented guidelines** which help public servants to navigate the specific challenges AI poses. While high-level principles are important, they are often too broad to guide decisions in practice, especially when those decisions require making tradeoffs [1], [2]. Often, **applications of AI will have the potential to both threaten and promote one or more of the principles**. For example, AI could enhance objectivity in decision-making by minimising human biases, but may introduce biases of its own, threatening the very same principle. We will also inevitably encounter tensions between the benefits of using AI to substantially improve public services, and concerns about the transparency and accountability of such systems. High-level principles do not in themselves provide guidance for how to navigate such tensions.

3. We argue that one key such guideline should be: "**Do not automate value judgements**": AI systems should not be allowed to make or substantially contribute to decisions which rely on potentially controversial value judgements without human oversight. All value judgements relevant to the decisions of AI systems should be

made **explicit** and, as far as possible, be **validated** by relevant human decision-makers.

4. In this evidence, we begin by explaining how uses of AI in the public sector may raise new challenges to Objectivity, Openness, Accountability, and Leadership. We then outline how the guideline, "Do not automate value judgements" may help public servants to uphold these four standards as AI is increasingly used in decision-making.

How AI impacts the Principles of Public Life

Impacts of AI on Objectivity

5. Prominent in discussion of the ethics of AI is the concern that AI systems can become biased and discriminate against some individuals or social groups. For instance, the COMPAS algorithm for estimating recidivism (the risk of a defendant reoffending), widely used to inform bail decisions in the US, has been criticised for having a systematically higher false positive rate for black defendants than white defendants [3]. That is, black defendants were incorrectly classified as 'high risk' twice as often as white defendants. The potential for this type of 'algorithmic bias' directly threatens the principle of Objectivity.

6. However, the threat of algorithmic bias has to be weighed against the potential to better counteract existing human and social biases. Human decision making processes are known to be subject to a number of biases: tending to overweight immediate rewards and neglect longer-term considerations; being easily influenced by subtle 'framing effects', and seeing patterns where none exist, to name but a few [4]. AI systems have the potential to help reduce existing biases in the decision-making processes within the provision and delivery of public services, either by replacing (parts of) human decision making with automated systems or by helping to structure decision-making processes and provide supporting information in ways that can be shown to minimise biases. Many limitations of human reasoning occur simply because the human brain cannot process large amounts of information and so uses 'heuristics' as shortcuts [5]; using AI to more effectively filter and process information could therefore help overcome some of these limitations.

7. The threat of algorithmic bias also has to be weighed against other values: it is not always possible to eliminate all biases and at the same time retain other desirable features of an automated decision system. For instance, in the case of the COMPAS recidivism algorithm, the underlying data on re-arrest rates made it mathematically impossible to avoid unequal false positive rates while preserving 'calibration within groups', i.e., the requirement that a given risk-score corresponds to the same level of

recidivism for all individuals [6]. If this requirement is not satisfied, the risk scores of the system would no longer mean the same thing for different groups (e.g. a score of 'high' would correspond to a lower risk for some groups than others). Arguably, this would itself compromise the objectivity of the risk scores.

Impacts of AI on Openness

8. AI systems can impact the openness and transparency of decision making in three key ways: (1) if the developers or deployers of a system decide to withhold information about how a system works (lack of **accessibility**); (2) if a system uses complex decision-making processes that even a human expert cannot understand (lack of **intelligibility**); (3) or if we cannot fully explain why a decision made by an AI system is accurate or adequate (lack of **explainability**). [7]

9. Lack of **accessibility** can be addressed by implementing appropriate methods to ensure the public, regulators, or auditing bodies can gain access to the decision processes of AI systems. In some cases, this might allow AI systems to be more transparent than many of the decision processes that occur within a human mind or complex bureaucracy.

10. There is ongoing technical work which seeks to address the challenge of **intelligibility** by developing automated tools for extracting relevant information about how an AI system works [8], [9]. It remains an open question to what extent these technical tools can fully address the challenge: there are upper limits to the level of complexity that can be fully comprehended by the human mind.

11. **Explainability** requires the ability to link our understanding of how the system works to other parts of our knowledge. We can distinguish between two types of explainability required to assess the adequacy of a decision: **causal explainability** and **value explainability**.

12. Causal explainability links the behaviour of an AI system to knowledge of the causal structures that underpin the accuracy and robustness of a given decision process [10]. For instance, if a predictive policing system uses the amount of litter on a street to predict the likelihood of crimes taking place, explainability concerns our ability to explain why and under what circumstances litter reliably correlates with crime.

13. Value explainability requires an understanding of the value judgements that underpin the adequacy of a decision. For instance, a predictive policing system may be designed to indicate in which part of a city policing is likely to lead to the largest number of arrests which lead to successful prosecutions. If it is not further constrained, such a system may end up prioritising street crime over financial fraud

(e.g. because the latter is more difficult to detect and prosecute). The adequacy of this decision implicitly relies (in the sense defined below) on the value judgement that de-prioritising fraud is an acceptable means of increasing the number of successfully prosecuted crimes. However, if the system simply gives an output in terms of which parts of the city police resources should be directed, this value judgement is rendered opaque.

14. The power of advanced machine learning relies exactly on its potential to discover correlations that are currently unknown by any human. This means that even if the relevant correlations that the system relies on could be made fully *accessible* and *intelligible*, we may still lack the pre-existing knowledge (i.e. of relevant causal structures and value judgements impacted) to make those correlations *explainable*.

15. Most contemporary machine learning systems are only designed to find and exploit correlations between input variables. Thus, they do not contain either information about causes or the value judgements impacted by its decisions. Unless the system is designed to discover causes or relevant value judgements, technical tools cannot extract the relevant information to make it explainable.

<u>Impacts of AI on Accountability and Leadership</u>

16. Both accountability for decisions and leadership in upholding public standards can be undermined by the use of opaque, automated systems in public sector decision-making. How can public servants take responsibility for their decisions if they are not fully in control of and do not fully understand the processes by which these decisions are made? How can public servants exhibit these principles in their own decisions when those decisions are taken out of their hand?

Why value judgments should not be automated

17. We have outlined just a few ways in which the use of AI in the public sector raises new challenges for upholding the Principles of Public Life; others will undoubtedly highlight more. We believe that proposing new principles or modifying existing ones is unlikely to be effective, since it would not help public servants to navigate the tensions that AI introduces within and between existing principles, or to make trade-offs in a given context.

18. We suggest that the Committee for Standards in Public Life should consider developing more specific guidelines for how to ensure the Principles are upheld within the specific context of using AI and automated decision-making. We suggest one such guideline: “Do not automate value judgements”. In this section, we outline what this guideline means in more detail and why we believe it is important.

What are value judgements?

19. By value judgements, we mean **concrete, practical judgements of the ethical worth** of a given course of action. While some philosophical theories claim, in principle, to be able to answer all ethical questions, in practice these theories cannot be implemented mechanically like formulae, but require humans on the ground to make judgements [11]. Almost all ethicists agree that deciding on an ethical course of action in practice requires **knowledge of the context**. This includes assessing (a) which ***ends*** we should aim to achieve, (b) what are acceptable ***means*** to achieve those ends, and (c) what the appropriate way is to ***balance*** between competing moral considerations.

20. For example, in healthcare provision, (a) one valuable *end* is to achieve the best overall health-outcomes from the available resources. However, it would not be acceptable to achieve this by (b) *means* of removing all care from patients less likely to recover, even if this could be shown to maximise overall health outcomes. On the other hand, some prioritisation of resources is necessary. Exactly how to (c) *balance* efficiency and equity of healthcare provision in concrete cases is an example of a value judgement.

21. One complicating factor is that value judgements might be implicit in a system. For example, if no one has thought to constrain a system from doing something, then implicitly that system has been permitted to do it. To extend the example above, an AI system programmed to optimise overall health outcomes within a hospital might determine the verdict that this is best achieved by redistributing resources from the seriously ill to more effectively cure those with milder afflictions. The system was not explicitly programmed with the value judgment that this would be an acceptable means to achieve the goal of optimising health outcomes. Nonetheless, its decision can still be said to implicitly rely on this value judgement, since its decisions will only be acceptable to those who consider that this would be an acceptable means.

22. In a given community, people's value judgements on a given issue can vary widely. It is a basic fact of both academic and public ethical discourse that reasonable people disagree on what constitutes ethically desirable and acceptable action. It is one of the core functions of public institutions (such as voting, public consultation, courts, etc.) to identify and sanction value judgements which individuals can accept as a legitimate basis for collective action, even if they disagree.

23. Holders of public office are either directly part of these institutions or are charged with carrying out their decisions. They have a duty to ensure that the provision and

delivery of public service embodies legitimate and appropriately sanctioned value judgements.

Why value judgements should not be automated

24. Ensuring that value judgements remain a human responsibility, and are made as transparent as possible, would promote the principles of public life in the following ways:

25. Objectivity: Algorithmic bias poses a threat to the objectivity of decision-making in the public sector. We do not propose a general principle for deciding whether a given type of bias is an acceptable price for maintaining other desirable requirements, or for how to compare the potential benefits of reducing human biases against the costs of introducing new ones. These questions will rely on contextual value judgements. **What is crucial is that these value judgements are made explicit and transparent, so that they are open to public scrutiny and legitimation through appropriate institutions.** For instance, the deeper problem in the COMPAS case was that even if maintaining the 'calibration within groups' requirement could be argued to justify the resulting biases, this value judgement was not made explicit and transparent until critical data audits by investigative journalists brought it to public attention.

26. Openness: The fact that we cannot always explain how an AI system made a decision and whether that process was adequate challenges public servants' ability to make decisions in an open and transparent manner. **In the absence of full causal explainability, value explainability becomes all the more important.** Value explainability can compensate for causal explainability. Causal explainability is useful for debugging AI systems and predicting when they will behave reliably. However, this is only a means to achieving the objectives that are the subject of value judgements. Thus, if we can be fully explicit and transparent about which value judgements a system relies on, and validate that those value judgements are indeed respected, causal explainability becomes irrelevant. [12]

27. Full value transparency requires at the very least making it transparent: **(a) which *ends* the system is designed to optimise, (b)** what constraints it is designed to respect in terms of **the *means* by which it achieves those ends** and **(c) what trade-offs between different values it makes**, including how well it performs for different types of predictions (e.g. men vs. women, street crime vs. fraud, etc.). Full value transparency may be easier to achieve than full causal explainability, since it depends in large part on the judgements of designers and developers.

28. Accountability and Leadership: Value transparency also allows public servants to remain accountable even when decisions are automated, by **enabling others to validate that the values pursued are in fact ones that are legitimated by appropriate democratic processes**.

29. The democratic process of identifying and agreeing on appropriate values in itself contributes to the legitimation of value choices. By insisting that these remain transparent and not (fully) automated, we protect the integrity of these democratic processes.

30. Finally, value transparency ensures that leadership can be maintained**,** by **making holders of public office responsible for identifying which values** a system ought to pursue and monitoring that it does so adequately. This is relevant both in the procurement and the deployment of AI systems.

References

[1] Whittlestone, J. Nyrup, R. Alexandrova, A. Dihal, K. Cave, S. (2019) Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research. London: Nuffield Foundation.
[2] Whittlestone, J., Nyrup, R., Alexandrova, A., & Cave, S. (2019, February). The Role and Limits of Principles in AI Ethics: Towards a Focus on Tensions. In Proceedings of the 2nd AAAI/ACM Conference on AI, Ethics, and Society. AAAI and ACM Digital Libraries.
[3] Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias. ProPublica, May, 23.
[4] Kahneman, D. (2011). Thinking, fast and slow. Macmillan.
[5] Lieder, F., & Griffiths, T. (2019). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. Behavioral and Brain Sciences, 1-85.
[6] Kleinberg, J., Mullainathan, S., & Raghavan, M. (2016). Inherent trade-offs in the fair determination of risk scores. arXiv preprint arXiv:1609.05807.
[7] Selbst, A. D., & Barocas, S. (2018). The intuitive appeal of explainable machines. Fordham L. Rev., 87, 1085.
[8] Biran, O. & Cotton, C. (2017). Explanation and justification in machine learning: a survey. *IJCAI 2017 Workshop on Explainable Artificial Intelligence (XAI).* http://home.earthlink.net/~dwaha/research/meetings/ijcai17-xai/
[9] Lipton, Z. (2017). The mythos of model interpretability. *2016 ICML Workshop on Human Interpretability in Machine Learning (WHI 2016)*. https://arxiv.org/abs/1606.03490
[10] Caruana et al. 2015. Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1721-1730.
[11] Brundage, M. (2014). Limitations and risks of machine ethics. Journal of Experimental & Theoretical Artificial Intelligence, 26(3), 355-372.
[12] Kroll. 2018. The Fallacy of Inscrutability. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 376. https://doi.org/10.1098/rsta.2018.0084

[redacted] [redacted]

[redacted]

SUBMISSION 10

[redacted]

[redacted] [redacted]

Dear CSPL,

On a personal basis, I'm making a short submission to your review of AI and Public Standards.

I believe that the Nolan Principles stand the test of time, but the domain of their application is shifting. With increased automation and the future use of AI/ML, decision-making relevant to the Nolan Principles will be ever more deeply embedded within operational contexts. I am unconvinced that current governance protocols are sufficiently geared towards surfacing risks, issues and performance commensurate with these new demands.

While the Nolan Principles apply to all aspects of the public sector, they are not reflected sufficiently in the domain where AI/ML decisions will be taken. For example, here is the Government Functional Standard, GovS 002: Project delivery document:

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/746400/Project_Delivery_Standard_1.2.pdf

In this document there are important principles stated, and there are areas such as assurance and risk management that could encompass a great deal. However, there is scope to incorporate / embed the Nolan Principles more clearly and explicitly, and this could lead to a greater focus on issues relevant to standards in public life where decisions are made.

Historically, such determinations may be seen as the preserve of policy decisions, but this is affected by two factors. Firstly, with the added complexities of automation it is necessary to actively seek out standards-related issues as they arise rather than assume they are dealt with in the formation of the policy. Secondly, agile policy-making (and use of agile methods in delivery) requires robust surfacing of issues from the front line of operational decisions.

Perhaps the most important aspect of how the Nolan Principles impacts on this environment (beyond simple awareness of those involved) is to do with monitoring and transparency. Therefore, operational guidance such as this should seek to expand its scope beyond value for money and timeliness and incorporate:

- Nolan-principle-based risk/impact assessment and management
- More robust requirements for impact monitoring
- Specific requirements for registration to canonical cross-government lists of use of personal data, algorithimic decision-making of all kinds, especially AI/ML

Without these kinds of controls more robustly and explicitly in place it will be hard for Ministers and senior officials to, hand on heart, commit that they are meeting the Nolan Principles when they agree to deploy automation that affects the general public in any substantial way.

Kind regards

David Evans

[redacted]

SUBMISSION 11

BCS, The Chartered Institute for IT - Consultation Response to:

Artificial Intelligence (AI) and public standards review, July 2019

BCS

The Chartered Inst tute for IT

F rst F oor, B ock D

North Star House

North Star Avenue

Sw ndon SN2 1FA

BCS, The Chartered Institute for IT

The purpose of BCS as defined by its Royal Charter is to promote and advance the education and practice of computing for the benefit of the public. We bring together industry, academics, practitioners and government to share knowledge, promote new thinking, inform the design of new curricula, shape public policy and inform the public. As the professional membership and accreditation body for IT, we serve over 60,000 members including practitioners, businesses, academics and students, in the UK and internationally. We also accredit the computing degree courses in ninety eight universities around the UK. As a leading IT qualification body, we offer a range of widely recognised professional and end-user qualifications.

This document is the BCS response to the Public Standards Committee review[8] into artificial intelligence (AI) and standards.

Main Conclusion

Machine Learning often seems closed to scrutiny, unfathomable in the way it draws conclusions and often appears beyond human comprehension. What it actually does is uncover patterns in data through an iterative process of approximation, to gradually build better statistical data models relating to the real world. These patters should be thought of as the 'best guess' possible from the current data, which become gradually more accurate as more data becomes available.

From a public service context, it therefore seems helpful to view AI as a new form of statistical-inference technology, which although potentially incredibly powerful should be seen as an incremental advance on existing data-driven, statistical technologies needing appropriate ethical checks and balances, especially when drawing conclusions relating to public policy. Given this context, we believe it is perfectly possible to update existing standards to ensure they are applicable to AI enabled public services.

We believe that to uphold the principles of public service we should treat AI not as Artificial Intelligence, but as Artificial Intuition. By framing AI as a form of computational intuition we would be far more circumspect in our reliance on and trust in it. We would probably be more conscientious about building appropriate safeguards to make sure it genuinely enhances standards of public services.

There has been a significant amount of work over recent years into developing guidance, frameworks and standards for the adoption of data science methodologies, techniques and applications in public service that are highly relevant to the possible future adoption of AI. In our view, given sufficient resource and coordination across relevant stakeholders these can be readily updated to take into account any new concerns arising from the adoption of AI.

[8] AI and Public Standards – Terms of Reference - GOV.UK

In 2015 the Public Standards committee published guidance[9] on Ethical Standards for Providers of Public Services. This was intended to provide:

> "*practical guidance to both providers of public services in building and embedding ethical standards in an organisation, and to commissioners in setting ethical expectations for the delivery of public services as well as ensuring those standards are met*".

In our view that guidance is relevant to commercial providers of AI enabled public services commissioned by government, with some potential updating to ensure it takes account of new concerns that are particular to AI.

Government has developed extensive public service standards[10] for the development of online services in recent years. Government has also published a data ethics framework[11], which includes recommendations covering adoption of Machine Learning techniques and other Data Science recommendations.

Recently the IEEE have published[12] a comprehensive range of AI related high-level ethical principles, key issues, and practical recommendations that should prove to be useful in updating public service standards that apply to software engineers, data scientists and AI practitioners.

Lastly, the Singapore Government recently published[13] a draft Model AI Governance Framework aimed at commercial companies creating AI enhanced products or services, which we feel is likely to be relevant and applicable to the UK.
We feel that the above guidance, standards and frameworks provide a comprehensive base from which to further develop applicable updated standards for AI enhanced public services.

1.a Current use of artificial intelligence and associated advanced technologies in the public sector

There is a very old adage in computer science that sums up many of the concerns around AI enabled public services:

> "*Garbage in, garbage out*"

In other words, if you put poor, partial, flawed data into a computer it will mindlessly follow its programming and output poor, partial, flawed computations. AI is a statistical-inference technology that learns by example. This means if we allow AI systems to learn from 'garbage' examples, then we will end up with a statistical-inference model that is really good at producing 'garbage' inferences. Worse, once an AI system has assimilated garbage into its inference model, it will produce garbage results even from good quality inputs.

---

9 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/481535/6.1291_CO_LAL_Ethical_standards_of_public_life_report_Interactive__2_.pdf

10 https://www.gov.uk/service-manual/service-standard

11 DCMS Data_Ethics_Framework.pdf

12 https://standards.ieee.org/industry-connections/ec/autonomous-systems.html

13 A Proposed Model AI Governance Framework

There appear to be too many cases of poorly thought through, rushed, and badly implemented AI enhanced technologies that are causing harm due to the 'garbage in, garbage out' problem.

In March 2019, New York University academics published[14] a research paper[15] showing how 'dirty data' have invalidated some predictive policing systems, which is an example of 'garbage in, garbage out'. It transpires that some data used to develop Machine Learning models for some predictive policing applications were invalid, since they were based on criminally corrupt activities of some rogue police officers. This invalid data is termed 'dirty data'. This example highlights the need for ensuring the provenance of data is very carefully validated before it is used for a system that has the potential to cause significant harm and significant waste of public money.

Healthcare is an area where AI technology is allowing cheap access to commercial personal medical diagnostic services that were previously unaffordable to many, and until now most people have only been able to access through the public sector. The risk with such new services is that rushed, and poorly developed healthcare products could make false claims that seriously undermine public trust in such technologies. Alarmingly, a recent article in the Lancet[16] raised significant questions about the medical competency of some healthcare apps and raised questions about their potentially bogus marketing claims. The comments in the Lancet article suggest that some healthcare apps currently available are suffering from the 'garbage in, garbage out' issue. The public should have complete trust in the medical expertise embedded in such apps and should not have to scour medical journals such as the Lancet to determine their validity.

1.b How standards may be affected by the widespread introduction of these technologies into the public sector

In July 2019 the University of Essex published a report[17] that found there have been 'significant flaws' in the way UK police forces have trialled AI enabled facial recognition technology. Apart from 'garbage in, garbage out' issues that appear to have resulted in the facial recognition technology often falsely identifying someone as a criminal suspect, the research also identified operational issues including:

- inconsistencies in the process of officers verifying a match made by the technology;
- a presumption to intervene when a suspect was identified;
- how the police engaged with individuals;
- and difficulties in defining and obtaining consent of those affected.

These issues show the potential for harm when flawed statistical-inference is combined with operational procedures that fall short of public service standards. This example shows an understandable tendency for public bodies to presume AI enabled technology is infallible, which may be due to the false presumption that they actually are intelligent. If AI stood for

[14] https://www.nyu.edu/about/news-publications/news/2019/march/predictive-policing-is-tainted-by--dirty-data---study-finds.html
[15] https://www.nyulawreview.org/wp-content/uploads/2019/04/NYULawReview-94-Richardson-Schultz-Crawford.pdf
[16] https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(18)32819-8/fulltext
[17] https://www.essex.ac.uk/news/2019/07/03/met-police-live-facial-recognition-trial-concerns

‘Artificial Intuition’ rather than ‘Artificial Intelligence’ hopefully humans would react differently to outputs from those systems.

To summarise, there appears to be a significant issue when AI systems are assumed to be absolutely correct in all their inferences, when they should be thought of as outputting a best guess based on patterns learnt from a set of examples, some of which may be invalid and that may not be sufficiently representative. Treating AI as a statistical-inference technology in frameworks and standards may help mitigate against such issues.

1.c What safeguards and considerations of standards are currently in place in technology procurement processes in the public sector

Please see our main conclusions for a list of standards currently in place that we believe are relevant to AI enhanced public services.

To understand what safeguards may be necessary it is helpful to understand what might cause AI enabled services to be flawed. From consultations we have held over the last eighteen months with academics and employers we have found significant anxiety in the IT profession and beyond about potential misuse of data leading to poor standards in AI enabled services. For example, the follow list illustrates how unprofessional practice can lead to poor standards of service:

- poor data governance resulting in companies unable to properly monitor how data is used, who is using the data, or where duplicates of data is stored, which may result in unethical practice going undetected
- lack of diversity in product development teams, which is a concern as non-diverse teams may be more likely to follow practices that inadvertently hard-wire bias into new products or services
- using incomplete data to incorrectly infer personal characteristics,
- allowing data to be improperly shared within an organisation
- improperly aggregating data from different sources to infer personal characteristics, and without making it obvious that such aggregation is occurring or for what purpose
- incorrectly cleaning data, leading to invalid data
- incorrectly disambiguating data resulting in the wrong data being associated with an individual
- incorrectly merging different data pipelines from third parties
- not conducting proper due diligence to ensure correct provenance of data through the supply chain (which may well be offshored and distributed across different national jurisdictions)
- using data analysis methodologies that are invalid in a particular context
- applying analytical models as part of decision making processes that are poorly tested (including for example inappropriate Machine Learning based neural networks)
- using invalid anonymisation techniques that do not provide sufficient protection against deanonymisation
- storing data insecurely so that it is at risk of being misappropriated by bad actors

These issues are mainly around poor data engineering and poor data science methodologies, rather than specific to AI. Hence if AI enhanced public service providers were obliged to follow the DCSM Data Ethics Framework, most of these issues would not

occur. If procurement standards followed the 2015 guidance from the Public Standards committee refenced earlier, incorporating the DCMS Data Ethics Framework this would make for robust procurement standards. Treating AI as a statistical-inference technology within procurement would allow relatively straightforward adaptation of the guidance and framework so that it could be applied to AI enhanced services.

1.d What safeguards and considerations of standards are currently in place in the deployment of AI and advanced technologies within the public sector

Please see our answer to 1.c, which also covers this question.

1.e What safeguards and considerations of standards are currently in place in private sector organisations developing AI services intended for use in the public sector

In our view there is great variation in standards across private sector organisations, some of which are exemplary whilst others have work to do. BCS is concerned that this significant variation in standards and potential for poor ethical practice may undermine the industry's public and political credibility and consequently its capacity for innovation and commercial viability.

In autumn 2018, BCS surveyed its members on topics including ethical standards across the IT industry, receiving responses from over 4000 members, ranging from UK students in higher education to Chartered members and senior Fellows. Figure 1 shows the result of one of the questions from the survey that asked members for their perception of ethical standards in the IT sector. The bottom line of the table shows the aggregation of the scores for the 'Neither high nor low', 'Quite low' and 'Very Low' categories. Collectively those three categories are termed 'not high'. What this table shows is that 47% of UK students chose 'not high' categories for ethical standards, compared to over 64% of Chartered IT Professionals and Fellows who chose 'not high' categories.

Current State of Ethical Standard	Overall	Fellows	MBCS CITP	MBCS	AMBCS	Affiliate	Student UK	Student Non-UK
Neither high nor low	36%	35%	38%	38%	33%	33%	25%	33%
Quite high	28%	24%	26%	28%	34%	20%	29%	25%
Quite low	20%	28%	22%	20%	16%	20%	17%	15%
Don't know	8%	6%	7%	7%	10%	16%	12%	9%
Very high	4%	1%	3%	3%	5%	8%	12%	14%
Very low	4%	5%	4%	4%	1%	3%	5%	4%
Sum of not high categories	**60%**	**68%**	**64%**	**62%**	**50%**	**56%**	**47%**	**52%**

Figure 1

Overall across all categories of BCS members 60% chose a 'not high' category of ethical standard. In other words, 60% of members appear to believe there is significant room for improvement in the ethical practices they observe.

In contrast when we asked BCS members how important ethical practice was to them, 84% of them responded that ethical practice was very important, as shown in Figure 2. This suggests the underlying issues in organisational trust may have more to do with how well organisations institutionalise ethical practices, rather than any reluctance on the part of IT professionals to adopt ethical practices.

Importance of ethical practice	Overall	Fellows	MBCS CITP	MBCS	AMBCS	Affiliate	Student UK	Student Non-UK
Very important	84%	86%	86%	84%	83%	88%	80%	80%
Quite important	13%	9%	11%	13%	12%	10%	12%	19%
Not very important	2%	3%	2%	2%	3%	2%	3%	1%
No opinion	1%	1%	1%	1%	2%	0%	0%	0%
Not at all important	0%	1%	0%	0%	1%	0%	5%	0%

Figure 2

2. Examine how provisions for standards can be built into the development, commissioning, and deployment of new technologies in the public sector

In our view an important purpose of standards is to ensure companies can demonstrate they are both competent and ethical. Our work with the Office for AI consulting widely with employers highlighted that their most urgent business need is for **diverse interdisciplinary** teams able to **ethically**:

- transfer **scientific** knowledge into business contexts
- **engineer** systems that meet strategic business objectives
- **manage** the adoption of technology and maximise its value across strategic business units

BCS feels it is therefore important that companies developing or deploying AI-enhanced public services must demonstrate they have the skills to address the above needs. Working with a range of employers we've identified a high-level set of skills that broadly encompass what it means to be a competent and ethical AI practitioners. These are:

Science skills for Artificial Intelligence Professionals

An AI Practitioner:

- Understands how AI and particularly Machine Learning algorithms, techniques and methodologies are designed, developed, optimised and applied at scale to achieve business objectives.
- Can select and use appropriate statistical methods for sampling, distribution assessment, bias and error.
- Understands AI and particularly Machine Learning problem structuring methods and can evaluate which method is most appropriate for business needs.

- Applies rigorous scientific methodologies through experimental design, exploratory modelling and hypothesis testing to reach robust conclusions, and can explain how those are reached to internal and external stakeholders.

Data engineering skills for Artificial Intelligence Professionals

An AI Practitioner:

- Has a demonstrable understanding of how to expose data from systems, how to efficiently extract data from potentially heterogeneous source systems, and how to ensure standards of data quality and consistency for processing by AI systems.
- Works with other technologists and analysts to integrate separate data sources in order to map, produce, transform and test new scalable AI products and services that meet user needs.
- Works with other technologists and analysts to understand and make use of different types of data models.
- Understands and can make use of different data engineering tools for repeatable data processing and is able to compare between different data models.
- Understands how to build scalable machine learning pipelines and combine feature engineering with optimisation methods to improve the data product performance.

Product development skills for Artificial Intelligence Professionals

An AI Practitioner:

- Uses a range of professional coding practices to build reliable, reusable, scalable AI products and services to time, quality and budget
- Can work as part of a team to effectively integrate AI technologies into business systems.
- Can take into account non-functional requirements such as system performance and integration requirements as part of an enterprise systems perspective
- Understands the enabling infrastructure required to support AI technologies
- Can demonstrate why AI products and services are valid against user requirements in a manner comprehensible to the relevant internal and external stakeholders.
- Works in accordance with agreed software development standards, including security, accessibility and version control.

Business skills for Artificial Intelligence Professionals

An AI Practitioner:

- Understands the context of the business including its processes, data, priorities and its wider values, objectives and strategy.
- Works collaboratively with domain experts to fully understand the requirements, checking understanding and testing models and solutions throughout the engagement
- Can effectively communicate the value, opportunities and limitations of AI technologies to a range of audiences with varying technical background.
- Uses the most appropriate medium to visualise AI based outputs to tell compelling and actionable stories relevant for business goals.

- Maintains a user focus to design AI solutions that meet user needs, taking account of ethical issues.
- Is familiar with the state of the art of techniques that help in modelling and understanding a business and its operation.
- As part of a team,
  - is able to support the scoping and business priority setting for large or complex changes caused by the adoption of AI, engaging senior stakeholders as required
  - is able to help identify the impact of adopting AI on business value and performance
  - uses the appropriate methods and techniques for the assessment and management of business risks that might result from adopting AI technologies

Ethical concerns for Artificial Intelligence Professionals

An AI practitioner is aware of and considers ethical concerns relating to the design, development, deployment, management and maintenance of AI products and services, such as for example

- Unfair or prejudiced bias in data or models
- Potential unconscious bias of AI practitioners and product development teams
- Appropriate level of transparency in design and development of AI models
- The impact of AI on restricting or enhancing user autonomy and wellbeing, whether in the workforce, in the customer base or society at large
- The ability of individuals to have appropriate control over their personal data
- Potential unintended, inappropriate or malicious use of AI products or services

The AI practitioner proactively works with and supports organisational stakeholders to develop appropriate policies, processes and practices to prevent unethical issues arising from the design, development, deployment, or management of AI products and services. They also proactively support their organisation to improve the ethnic and gender diversity and inclusivity of the workforce at all levels.

References

Much of the above draft framework has been adapted or synthesised from these sources:

- Digital, Data and Technology Profession Capability Framework[18]
  - Data scientist: skills they need
  - Data Engineer: skills they need
- Skills Framework for the Information Age[19]
  - Skill: Solution architecture - Level 5
  - Skill: Data management - Level 5
  - Skill: Business analysis - Level 5
  - Skill: Business modelling - Level 5

[18] https://www.gov.uk/government/collections/digital-data-and-technology-profession-capability-framework

[19] https://www.sfia-online.org/en/framework/sfia-7/busskills/level-5

    - Skill: Requirements definition and management - Level 5
- Google's definition for Professional Data Engineer[20]
- The Life of a Data Engineer[21]

We feel when commissioning services from companies, those companies should be able to evidence their AI related workforce has the relevant skills based on the above. It is worth investigating whether it would be appropriate for procurement and commissioning standards to mandate that companies have competencies and ethical practices that are consistent with the above.

3. Consider to what extent the use of artificial intelligence and associated advanced technology has implications for our understanding and formulation of the Seven Principles of Public Life

Selflessness

AI can help us better understand what the public interest is, because it can inform our decisions based on analysis of vast data sets in ways that were not possible before. However, the limitations of AI as a statistical-inference technology means it can only ever supplement a range of traditional methods of understanding what is in the public interest.

Integrity

On July 10th 2019 the Government announced[22] Amazon's Alexa will be partnering with the NHS to allow people to ask Alexa to find health related advice from NHS sources. This example is seen by some as controversial although the intention is entirely to ensure people have access to trusted advice from an authoritative source. In part it is seen as controversial because although Amazon is a reputable and competent company, the public will have no way to verify whether or not Amazon act ethically. Members of the public might have doubts about how much they should rely on promises of confidentiality considering the recent discovery[23] that some Google workers in different parts of the world listen to around 0.2% of all voice recordings from its smart speakers.

The above example illustrates how easily good intentions can inadvertently lead to potential pubic concerns about integrity. It suggests more open and comprehensive dialogue with the public should be a precursor to allowing commercial interests to deploy public services.

Objectivity, Openness, Accountability

As highlighted by our answer to 1.b, judgement can be clouded by over-reliance on AI, which can lead to harm. AI over other advanced technologies can cloud our judgement due to the often 'black box' nature of its inferences. We feel so long as AI is limited to public services where it is used as a statistical-inference technology, and operates within technical boundaries where it can be constrained to be fit for purpose then it will hugely enhance public service benefit.

---

[20] https://cloud.google.com/certification/data-engineer
[21] https://www.mastersindatascience.org/careers/data-engineer/
[22] https://www.bbc.co.uk/news/health-48925345
[23] https://www.independent.co.uk/life-style/gadgets-and-tech/news/google-home-recordings-listen-privacy-amazon-alexa-hack-a9002096.html

AI – should we think of it as Artificial Intuition

Perhaps the best way to ensure the principles of public service are supported by and not weakened by AI is to regard it as a form of incredibly powerful Artificial Intuition.

SUBMISSION 12

Submission of evidence to Nolan Committee

Enquiry on AI and Public Standards

Dr Emma Carmel[1]
Associate Professor
Social and Policy Sciences
University of Bath

[redacted]

Date of submission: 30 July 2019

Automation of human decision-making through the use of artificial intelligence in government and public service has the potential to transform the nature of public service in a democratic society.

Its use transforms the relationship between state and citizen; among individual citizens; and among groups of citizens. These developments significantly challenge our systems of political responsibility and accountability, and create new roles for public servants.

As such the Nolan Committee's Enquiry into AI and pubic standards (henceforth 'the Enquiry'), is warmly welcome, timely, and should be the start of a wide-ranging and long-term discussion. This submission is intended to contribute to this discussion and is set out as follows:

Part I: Scope and definitions
Part II: Main submission
This is organised around the four topics in the Enquiry's terms of reference. The response to each topic includes key points, followed by evidence and argumentation. Topic 4 includes recommendations to the Enquiry.

Annex: Detailed considerations for provision for ethical standards in automated decision-making

[1] Dr Carmel researches how public policy is made and delivered, and the changing characteristics and power of 'the state'. In the last 20 years she has published widely on EU, UK and comparative governance of migration, social security and labour market policies, as well as governance theory. She was recently appointed pathway lead for *AI and Public Policy* at the UKRI-funded doctoral training centre in 'Accountable, Responsible and Transparent AI' at the University of Bath. All opinions expressed in this submission are her own.

Part I: Scope and definitions

Automated decision-making systems

Following the broad scope of the Enquiry, this submission addresses the implications of using automated decision-making systems (ADS) in public policy decision-making and service provision. ADS that use any AI technique represent the functional cutting edge of the application of artificial intelligence[2] to public policy.

Public policy ADS are a mix of conventional human decision-making, technological procedures and capacity, and their interaction.

In this submission 'public policy ADS' refers to all aspects of the following activities and processes. Unless otherwise specified, I refer to these activities collectively.

- decision to adopt ADS, including impact assessment and ethical review
- procedures for, and terms and conditions of procurement
- all model, algorithm, system design and iteration
- all data selection, cleaning, harmonisation, storage and formatting
- learning or automated system functioning, application and outcomes in specific service context
- evaluation and audit of system adoption, relevance, and use over time, in specific services and wider institutional and policy settings.
- system revision and termination

The *implications of ADS for the Seven Principles do not concern AI as a set of technologies.*

They concern the characteristics of these technologies <u>and</u> how they are embedded in the wider eco-system of decision-making and service provision. This includes public policy and services undertaken by, and in collaboration with, private sector actors.

AI and types of automation

Any decision or provision where regularised criteria lead to specific decision outcomes is 'automated'. This is a commonplace of public policy.

Indeed most public policy ADS currently in use[3] are, strictly speaking, systems for *<u>automated decision-recommendations</u>*. They offer risk-based recommendations for decisions, which use past data to assess likelihoods of future events. The decision to act on these recommendations in most cases remains in the hands of humans.

I have identified four types of automation, depending on the role of AI technologies in the ADS. These types pose different challenges for the Seven Principles and are discussed in the main submission.

[2] As defined in the Enquiry's terms of reference. This includes declarative programming, machine-learning, supervised, unsupervised and neural networks/deep learning.
[3] In 'predictive policing', social care placement prioritisation, visa application assessment, welfare provision.

Part II: Main submission

Topic 1 ADS in practice and existing frameworks

Key points

- Automated decisionmaking systems share features that change the conventional arrangements for policy and political responsibility.

- Existing ethical frameworks only partially address the challenges posed by these changes.

1. Ethical challenges of ADS in practice

The key challenges of using AI in decision-making are typically discussed as questions of 'explainability' and 'bias'.

These concepts are closely related to the Principles of openness, accountability and objectivity, but are inadequate to capture the full political, social and ethical implications of public policy ADS adoption, development and termination.

'Explainability', policy and political responsibility

ADS involve a step-change transformation in the practice of government, including digital government. They fundamentally alter where decision-making takes place, by whom and when. In so doing they transform political and institutional relationships in public policy and services.[4] In particular, they:

- Collapse the distinction between policy decisions, operational design and policy delivery that has been the hallmark of UK policymaking of the last half-century.[5]

- Close the gap between commissioners and providers of services.

- Promise accuracy, consistency and speed but risk reductionism, rigidity and opacity.

ADS *disorganises* existing ways of understanding and arranging responsibility for decision-making.[6] In doing so, it may threaten the application of the Principles of integrity and accountability in public services and public policy.

In ADS, political and policy responsibility are *distributed* across a complex field without a clear chain or hierarchy of responsibility, in which the roles of different institutions, public bodies, private providers of services and technologies becomes blurred.[7]

---

[4] Dunleavy, P., H. Margetts, S. Bastow, and J. Tinkler. 2006. *Digital Era Governance: IT Corporations, the State, and e-Government*. Oxford: Oxford University Press; Henman, P. 2010. *Governing Electronically: E-Government and the Reconfiguration of Public Administration, Policy and Power*. Basingstoke: Palgrave Macmillan.

[5] Carmel, E. ed., 2019. *Governance Analysis*. Edward Elgar.

[6] For example, this might typically be organized between a combination of legal-institutional (rule-based), bureaucratic (role-based) or market (contract-based) forms of accountability.

[7] Responsibility is *distributed* among technological elements, e.g. data labelling, training data selection, modelling decisions; and institutional elements, e.g. procurement, treatment of decision-recommendations, conditions for termination.

In this context, requirements for responsibility cannot be resolved by improving 'explainability' of algorithms or machine-learning, and the ways in which ADS and AI technologies intersect needs closer examination.

Types of automation

I have identified four ***types of automation*** in public policy ADS. Each has different, but related, implications for the practice and ethics of public policy and services.[8]

(1) *Single-stage algorithmic automation* can be based on pre-defined decision-models, organised around regulatory criteria for example. By itself, this is not strictly speaking 'AI'. An AI-based type (1) automation could, however, involve designing an algorithm to produce a particular outcome (decision), but not specify instructions of how to get that outcome.[9]

(2) *Two-stage algorithmic automation* uses machine-learning on data from previous decisions to generate a model of those decisions. This model is then used to design an algorithm that is applied in actual type (1) automation. Like all machine-learning (ML), this type of automation is in practice heavily dependant on the quality, detail and appropriateness of the data used for training.

(3) *Concurrent automation* uses machine-learning and/or neural networks in 'real-time' to make decision-recommendations. As such, the decision-recommendations use past and current data to inform the risk-based recommendation.[10]

(4) *Autonomisation* is where machine learning and neural network systems make decisions based on patterns in data, and these decisions have policy and legal effect. The use of type (4) in citizen-facing public policies has not yet been publically acknowledged, but would have the most potential significant implications for the application of the Seven Principles.[11]

'Bias' and 'explainability' in different types of automation

In types (1) and (2) automation, there is a significant reduction in the ability to process and respond to contextual, new, or unusual information. In particular:

- Tacit knowledge that is often important for human decision-making may not be recorded in training data, making any resulting decision-recommendation algorithm partial, unreliable, skewed, or just opaque.
- If new data becomes relevant, it cannot be included without re-running the ML training, and it may take considerable time before the data is extensive and reliable enough to do so.
- ML systems look for general patterns to generate models, but much policy, and case law, turns on the individual case and context. Citizen rights and entitlements cannot only be reduced to automated rules.[12]

[8] Discussed in topic 3, below.
[9] The former is 'imperative programming', of the classic step-by-step, 'if…then' type. The latter might use 'declarative programming' where the steps and the order to reach the required outcome are not pre-specified.
[10] Tazzioli, 2018. *op cit*. esp pp. 276-79.
[11] Most reference to 'AI' in public policy refers to types (2) and (3), and, sometimes erroneously, to type (1). In this submission, I primarily refer to types (2), (3) and (4) and their integration into public policy ADS.
[12] This is of course why many policies and service regulations allow for a measure of discretion in decisionmaking that has direct effects on citizens, communities and places.

- Not all state/citizen interaction can be turned into data for use in ML training. Many citizens – especially more vulnerable and less articulate – may need human advice, input and coaxing to provide information which is important to secure their rights.

In types (3) and (4) automation, prior decision-making outcomes become integral to future policy decisions. The ML product itself is producing policy and law. This is exacerbated as ML learns from the outcomes it generates itself, so that any errors in the learning can be reproduced and confirmed.[13] This is the origin of problems usually discussed as questions of 'bias' and 'explainability' in ADS.

However, bias is a feature of social and political relations and cannot resolved with a technological fix.[14] Indeed, the problem with ML-based ADS is not bias per se.

The risks of ML-based ADS result from flawed assumptions about social data, modelling, automation and decisions. It is when these assumptions are linked together in practice, that inequalities and errors in ADS are extended, intensified and automated. These risks challenge Principles of objectivity, openness and integrity.

In particular:

- Data is assumed to accurately represent a universe of reality, rather than a (partial) version of a limited reality.[15]
- This version of reality is then modelled, rather than applying authorised regulatory criteria
- Policy-irrelevant criteria may be used, or criteria may be weighted incorrectly, due to previous decision or training-data errors.
- Explaining how the technology adopted a decision-recommendation does not assign responsibility or accountability.

Public policy ADS are, therefore, not merely supplementary tools for policymaking or delivery, but have diverse, wide-ranging and potentially unpredictable implications for public accountability and citizens' lives.

2. Existing frameworks

The recent launch of a set of guidelines on use of AI in the public sector in the UK is welcome, but represent a nascent set of concerns rather than a fully-fledged approach[16]

---

[13] Sheehy, B., 2019. Algorithmic paranoia: the temporal governmentality of predictive policing', *Ethics and Information Technology*, 21:49–58

[14] Referring to 'bias' in public policy ADS also risks bias being reduced to protected characteristics. There may be other biases automated in such systems that are ethically, socially or politically undesirable that affect specific sub-groups of the population. See discussion in Topic 3.

[15] This is compounded when systems are also based on very limited and poor quality training, see e.g. Sànchez-Monedero, J. 2018. *The dataification of borders and management of refugees in Europe*, Cardiff University, https://datajusticeproject.net/wp-content/uploads/sites/30/2018/11/wp-refugees-borders.pdf. In general, administrative and government data is extensive and consistent. However, it is also significantly skewed by performance targets and funding priorities (that may not accord with relevant empirical developments, in crime, say). This problem is compounded by practical issues in the format, storage and labeling of data, all of which can affect what the ML system 'learns' and produces. boyd, d. & Crawford, K. 2012. Critical Questions for Big Data, *Information, Communication & Society*, 15(5), 662-679*;* Citron, D. K., & Pasquale, F. 2014. The scored society: Due process for automated predictions. *Washington Law Review*, 89(1), pp. 1–33; Dunleavy et al., op cit.

[16] https://www.gov.uk/government/collections/a-guide-to-using-artificial-intelligence-in-the-public-sector, last accessed 12 July 2019.

The UK government guidelines rely on the Turing Institute's framework, alongside the government's Data Ethics framework. The Institute's framework is careful, thorough and wide-ranging. Particularly welcome is its introduction of 'sustainability' as an ethical requirement for the adoption of AI systems.

However, this guide, and the government's accompanying documents, while informative, are voluntary and remain at a high level of abstraction. There is no single, explicit ethical or governance framework to secure the coherence of ethical review.[17]

The guidelines and advice are the shared responsibility of the Office for AI in BEIS, and the Government Digital Service. The OAI is also responsible for promoting the development of AI technologies and industries, and so has a conflicting interest, and the GDS has wide responsibilities to support digitalization of central government. It seems unlikely that either has the capacity or remit to supervise ethical review on the broader questions of ADS adoption and use in public policy.

In accordance with the Principles of integrity, honesty and leadership, ethical review should not be acceded to a government office housed in a Department under ministerial supervision.

Additional problems include:

- *Lack of clear institutional authority and ownership*
- *Overly general ethical framework*. It is not precise enough to identify sector-specific implications of ADS across the full range of public bodies
- *Over-emphasis on project orientation*. The ethical implications of ADS do not have a completion date, and should not be conceptualised as a project to be developed – and completed – like any other software or digital development
- *Overly centralised approach to the design of the general framework*. It does not provide tools to assist the full range of public policy actors and service providers in using AI-based ADS[18]
- *Overly de-centralised approach to implementation of the framework*. The development of governance frameworks is assigned to any number of individuals of varying seniority, with no explicit consideration of the overwhelming need for systematic consideration of a range of expertise[19]
- *Lack of specific and systematic protocols* for policymakers on what constitutes risks, merits and implications of ADS

[17] It will be difficult for policymakers, whose role is to navigate value conflicts and value priorities, to use this as a guide to everyday ethical practice. The very general framing of its ethical principles does not offer guidance on conflicts between the stated principles, nor in their application (e.g. it emphasizes 'social values' as if these are unproblematically shared and understood).

[18] Public bodies, local authorities, and contracted private providers of public services, are developing and using ADS, without a generally applicable or sector-specific framework or system for ethical review, and with only limited and haphazard access to necessary expertise.

[19] Whittaker, M., Crawford, K., Dobbe, R. et al 2018. *op cit.,* at pp. 32-40. In order to evaluate the design and implications of ADS, adopters should consider potentially relevant expertise from: direct stakeholders; data scientists; social scientists including statisticians and domain experts like ethnographers and qualitative researchers; technologists and computer scientists; clinicians; lawyers from diverse specialisms, and ethicists/philosophers. Current UK government guidelines only suggest using principles developed by the Turing Institute, and that these should be used by responsible project staff to develop their own governance framework. https://www.gov.uk/guidance/understanding-artificial-intelligence-ethics-and-safety.

- *Lack of specific and systematic procedures* for assessing the risks and merits of ADS in particular cases

Given the above, it is unclear how policymakers can have the confidence in their ethical practices and their ability to uphold the Principles in practice.

This applies to all public servants and private actors working on particular AI projects; specialists involved in shaping the adoption and function of ADS work in particular institutions or policy areas; and senior leaders responsible for securing a consistency of approach.

As a result, existing framework and guidelines are not adequate to meet the challenges posed by public policy ADS and related applications.

We require an urgent, long-term response to manage forthcoming developments.

Topic 2 How to provide for standards

Key points

- The need for specific, detailed ethical governance of AI in policy and public services, by the public and private sectors, is acute.
- There are alternative frameworks being developed, of which one of the most systematic and thorough is that developed by the Canadian federal government
- Key elements for provision of standards must include impact assessment; inter-disciplinary review; and protocols for AI use, both in general, and in particular policy areas.

Alternative approaches to AI adoption, use and termination

Public policy ADS have two characteristics that distinguish them as AI-based systems: the legal authority of public service providers, and the life-changing attributes of public policy.

This means that general ethics of AI guides and frameworks, which are proliferating, are not specific enough to apply to public policy ADS. There are alternatives, however.

The Canadian government has focused on both substantive and procedural protocols for the adoption, use and termination of AI in public policy. It has

- Developed clear substantive and procedural guidelines for procuring, adopting and using machine-learning systems in government.[20]
- Identified four distinct levels of impact of public policy ADS adoption.[21]
- At each level of impact, identified both the quality standards and the political approval level[22] required to meet ethical challenges posed by ADS.
- Rejected type (4) *autonomisation* for any ADS with the potential for high and/or enduring levels of impact.[23]

---

[20] Treasury Board, Directive on the Use of Machine Learning for Decision-Making, v. 2.7 https://docs.google.com/document/d/1LdciG-UYeokx3U7ZzRng3u4T3IHrBXXk9JddjjueQok/edit#, last accessed 24 May 2019. This Directive, and the algorithmic assessment processes were developed by an interdisciplinary group of public servants, with input from Departments and academics. It was led by the Treasury Board of Canada, in an open and publicly iterated process.

[21] The levels of impact are defined in the Directive, appendix B. They are distinguished by: **type** (health, rights, economic interests and sustainability); **extent** (breadth of impact across a number of citizens, communities or eco-systems); **intensity** (degree of impact on these individual and systems); **duration** (how long is the impact) and **reversibility** of impact. For example, level IV impact are 'very high' on all types, with wide extent, 'often leading to impacts that are irreversible and perpetual' (p. 10). On algorithmic impact assessments in general, see: Reisman, D., Schultz, J., Crawford, K. Whittaker, M. 2018. *Algorithmic impact assessments: a practical framework for public agency accountability*, AINow, April.

[22] https://canada-ca.github.io/digital-playbook-guide-numerique/views-vues/automated-decision-automatise/en/automated-decision.html. The standards of quality assurance, include requirements to involve inter-disciplinary expertise; open and high quality research basis for the technological elements of the ADS; access to, and review of, source code and data. These are relatively demanding, and they increase significantly as the assessed level of impact increases. The parameters are set out in the Directive on machine-learning, *op. cit.*, pp. 11-14.

[23] Levels III and IV require 'specific human intervention points during the decision-making process; and the final decision must be made by a human'. Treasury Board approval is necessary for level IV systems Treasury Board.

- Created an 'algorithmic impact assessment' process[24] to enable policymakers to consider the implications of adopting ADS and which level of impact applies in each case.
- An online questionnaire for public policymakers to use for their impact assessment.[25]

The Canadian system is very new, and will not be without flaws in practice.[26] However, it presents an alternative to the current voluntarist UK approach. It offers a comprehensive, systematic, and easy-to-use system, for policy decisionmakers who need protocols to assist them in making ethical, and very long-term decisions.

General considerations for how standards can 'be built into the development, commissioning and deployment' of new technologies are fourfold.

- *designation of effective standards require*: specific frameworks, both general and sector/policy-specific

- *rigorous and consistent application and use of standards requires*: that frameworks are translated into protocols, guidelines, and templates that encourage reflection and ethical skills development in policymakers

- *development of appropriate and feasible standards requires*: inter-disciplinary engagement in the design of frameworks, protocols and guidelines

- *application of comprehensive and systematic standards requires*: each stage of ADS design, adoption, development, iteration and termination to be included.

The annex of this submission sets out likely detailed considerations for provision for standards in AI-based ADS.

[24] https://canada-ca.github.io/digital-playbook-guide-numerique/views-vues/automated-decision-automatise/en/algorithmic-impact-assessment.html.
[25] https://canada-ca.github.io/aia-eia-js/
[26] For example, over time it could become overly bureaucratic, rigid and procedural, with policymakers focusing only on meeting the Directive and assessment requirements, rather than reflecting on sector- or context-specific ethical considerations. Or policymakers might focus too much on gaming the levels of impact in order to gain approval for a preferred system. Or they might undertake appropriate reflections and analysis at the approval-seeking stage, but once the system is in place, questions of efficiency and policy delivery take priority and the ethical functioning of the ADS is not subjected to ethical review. Reflecting on these risks informs the recommendations in topic 4 in this submission.

Topic 3 Implications for Seven Principles

Key points

- AI in public policy ADS poses challenges to the *application* of the Principles, rather than to their definition or scope.

- The Principles of *integrity, objectivity, accountability, openness* are particularly important in this context and should be directly addressed in any system of ethical regulation and review of AI in public policy

- There is a case for extending the Principles to include something like 'sustainability'

Reasoning on implications for Seven Principles

Consideration of the Principles has the important benefit of expanding the focus of ethical discussion away from 'bias' and 'explainability' in AI-based ADS.

Examining each Principle highlights both key ethical problems in ADS, as well as some of the practical problems in identifying how to apply the Principles in ADS.

Integrity

- AI-based ADS challenge public servants and private actors' ability to act with integrity, when determining actions that depend on close distinctions between inference and prediction; decision and decision-recommendation; responsibility and accountability.[27]
- All ADS raise questions of how 'front-line' staff can and should interpret the risk-based calculations or output decisions produced by the ADS, and the level of responsibility this assigns to them[28]
- Design of visual representations of uncertainty and calculation by software providers must also secure integrity of the ADS decision[29]

Objectivity

- Even type (1) automation may contain biases that are not evident, especially to technological specialists without knowledge of the policy field[30]

---

[27] Mackenzie A. 2015. The production of prediction: what does machine learning want? *European Journal of Cultural Studies* 18(4–5): 429–445

[28] Ananny M. 2016. Toward an ethics of algorithms: convening, observation, probability, and timeliness. *Science, Technology & Human Values* 41(1): 93–117; Esposito, E. 2015. Beyond the promise of security: Uncertainty as resource. *Telos, 170*, 89–107

[29] Coopmans C, Vertesi J, Lynch M, et al. (eds) *Representation in Scientific Practice Revisited*. Cambridge, MA: The MIT Press. See Henman, P. 1997. Computer technology. A political player in social policy processes, *Journal of Social Policy*, 26 (3), pp. 323–340; Tazzioli, M. 2018. Spy, track and archive: The temporality of visibility in Eurosur and Jora, *Security Dialogue*, 49(4) 272–288, p. 279; Wojciech S. et al., 2017. Evaluating the visualization of what a deep neural network has learned. *In IEEE Transactions on Neural Networks and Learning Systems*, 2017.

[30] Friedman, B., & Nissenbaum, H., 1996. Bias in computer systems. *ACM Transactions on Information Systems*, 14(3), 330–347; Zarsky, T., 2016. The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making. *Science, Technology, & Human Values*, 41(1), 118–132.

- There is a large risk that errors and biases in existing data and decision-making are intensified and reinforced into all decision-making over time[31]
- The origins of system bias may lie in previously systematically-biased human decision-making reflected in the data. They may also be introduced during mundane activities of data generation, cleaning and labelling[32]

Openness

- In automation types (2), (3) and (4), the opacity of machine-based inferences and the indeterminacy of patterns identified put the Principle of openness at risk[33]
- Based on patterns found in preceding data, these types are not be able to accommodate new contexts, regulatory or policy changes, however small
- Rigidities and biases are not always easy to predict, interpret or accommodate the complexity of the exceptional case, or the individual[34]
- This problem is more challenging in the case of minorities, including functional or unrecognised minorities who are not represented in the data[35]
- The collapse of policy, operations and delivery in ADS require corporate designers and data providers for ADS to be subject to rigorous conditions of openness.

Accountability

- The collapse of distinctions between policy, operational design and delivery signal a need to prioritise procedures and standards for accountability in ADS.
- The distributed character of policy responsibility in ADS means that having a 'human-in-the-loop' is not enough to secure accountability because decision origins will not be traceable nor easily understandable[36]

---

[31] Eubanks, V. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (New York); Amoore, L. & Hall, A. 2009. Taking people apart: digitised dissection and the body at the border, *Environment and Planning D*, 27(3): 444-64. Owen, T. 2015. *Crisis of the State in the Digital Age*, Oxford UP, pp. 168 ff. A problem of this type seems to underlie the significantly increased difficulties in securing visitor visas experienced by African scholars, artists and professionals to secure visitor visas for events in the UK. https://www.ft.com/content/0206dd56-87b0-11e9-a028-86cea8523dc2

[32] Richardson, R., Schultz, J. and Crawford, K. 2019. Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice. *New York University Law Review Online*. Retrieved from: https://www.nyulawreview.org/online-features/dirty-data-bad-predictions-how-civil-rights-violations-impact-police-data-predictive-policing-systems-and-justice/; Whittaker, M., Crawford, K., Dobbe, R. et al (2018) *AINow Report 2018*, AINow Institute, New York University, pp. 24-30.

[33] Ananny, M., & Crawford, K. 2018. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability, *New Media & Society*, *20*(3), 973–989. https://doi.org/10.1177/1461444816676645, p. 982; also: Crain M., 2018. The limits of transparency: data brokers and commodification. *New Media & Society* 20(1): 88–104.

[34] This appears to be the case in the deportation of some students caught up in the English-test cheating scandal recently investigated by the NAO. Molnar, P. and Gill, L. 2017. *Bots at the Gate. A human rights analysis of automated decision-making in Canada's immigration and refugee system* (University of Toronto), https://citizenlab.ca/wp-content/uploads/2018/09/IHRP-Automated-Systems-Report-Web-V2.pdf, p. 51.

[35] This could include minorities in the context of a particular policy (men over 85 in pension policy, for example; or lone parents in family policy) who may suffer systematic and automated deleterious effects unrelated to protected characteristics or policy intent. See, e.g., Ruppert, E. 2012. The Governmental Topologies of Database Devices. *Theory, Culture and Society*, 29(4-5), pp. 1-21

[36] This is one of the ways in which the gap between commissioning/providing services closes. The design of the machine-learning system, data preparation and selection, decision model, source code and algorithm all involve decisions about policy that shape outcomes in ways that are not easily predictable. Fink, K. 2018, Opening the government's black boxes: freedom of information and algorithmic accountability, *Information, Communication & Society*, 21:10, pp. 1453-1471; Henman, P. 2018. 'Of algorithms, apps and advice: digital social policy and service delivery', *Journal of Asian Public Policy*, DOI: 10.1080/17516234.2018.1495885; Molnar and Gill, op.cit., pp. 52-3.

- Nor is redress enough given the combination of speed and rigidity in decision-making that are possible in ADS[37]
- We need an ethical framework and protocols that enable policymakers and citizens to navigate the distribution of responsibility and accountability

Selflessness, *Leadership*, *Honesty*

These three Principles also have implications for how we address the risks and challenges of ADS, but they are more diffuse than the other four.

Key issues are for human decision-makers to

- exercise *leadership*, where necessary in delaying, halting or preventing the adoption/use of ADS that do not meet requisite ethical standards

- engage *selflessly* in ADS development and adoption, to focus on the long duration of ADS and their implications, rather than the (often high status and career-enhancing) project-based, exciting technological development

- engage *honestly* in ADS, acknowledging uncertainty, their lack of expertise, and the need for ethical self-reflection

Sustainability

There is a case to be made for extending the Seven Principles with an eighth that has particular resonance for ADS.[38]

All government decisions have far-reaching consequences of long duration, for both individuals and groups. The ethics of government requires that these individuals and groups, as well as Courts and historians, should in most circumstances be able to identify the evidence and reasoning on which decisions rest, sometimes long after those decisions were made.

- ADS involve dependence on very high levels of energy consumption, specific forms of hardware and digital storage.
- Significant ongoing investment and resources are required to maintain enduring access to ADS as they become 'legacy' systems.

ADS pose questions about how to manage the durability of decisions, with the impermanence of evidence, and the need to maintain systems for consistent, traceable decisions over time.

This would require public servants and all those in public life, to consider the *sustainability* of their actions.

[37] Edwards, L. and Veale, M. 2017. "Slave to the Algorithm? Why a 'Right to an Explanation' is Probably Not the Remedy You are Looking for," *16 Duke L. & Tech. Rev.* 18.; Molnar and Gill 2018, op cit.
[38] As proposed in the Turing Institute's guide.

Topic 4 Recommendations for regulatory change

Key points

- We need a new system of ethical governance of AI in public policy and services
- This system should be statutorily constituted with flexibility and the ability to review and adapt in response to political and scientific change
- This system should be led by a statutorily constituted arms-length public body, modelled on the Human Fertilisation and Embryology Authority (HFEA).[39]

Issues to consider in regulatory change

Any system to provide for ethical standards in the development and use of AI in public policy and services must meet three practical challenges:

- Scientific authority and capacity to respond to the speed and sophistication of technological developments
- Political authority and ethical sophistication to practically assess, and respond to, distributed responsibility in ADS
- Political and scientific authority and capacity to respond to the full range of policies and public bodies affected

The UK already has an extremely successful, globally recognised regulatory body in an unrelated but analogous field: the HFEA. Its particular strengths as a relevant model for ethical standards include:

- It works *across public and private sectors* to license the ethical conduct of medical and research activities in its field.
- It was established at a time of enormous and extremely rapid scientific and technological developments, in a field that is concerned with fundamental ethical concerns and rights.
- It has a statutory basis, but is arms-length from government and so not subject to regulatory capture.
- It is not funded by or governed with or by industry partners, but has a clear independent regulatory role
- It operates a widely-respected system of sophisticated ethical review in complex scientific, social and political contexts
- It has become a leader in ethics in its field.
- The HFEA provides a model that fits well with the UK's existing policy infrastructure.

Such an Authority could have a strong role, to licence AI in particular ADS (like the HFEA), or a weaker role of leading on standards, review, assessment and acting as ombudsman.

[39] https://www.hfea.gov.uk/about-us/

In any case, there are two specifications for such an Authority in the field of public policy and services.

First, any general system of ethical regulation for AI in public policy and service requires sectoral-based ethics review as well, to account for the specific risks and attributes of decision-making and its automation in particular policy areas. The most obvious of these are defence and security; social welfare and care; health; criminal justice and policing; immigration.

Second, because such an Authority would have oversight of activities of government, it would be preferable to have it report to parliament.

Recommendations

1. Create a statutorily constituted arms-length public body, modelled on the HFEA

2. The Authority should have a wide remit that applies to central and local government activities, all public bodies and any private sector actors involved in any aspect of public policy ADS with direct effect on citizens.

3. The Authority should have inter-disciplinary working board and staff with strong citizen representation.[40]

4. The Authority should be accountable to parliament

5. The Authority should

    a. licence and oversee the adoption of any AI system and public policy ADS by any policy actor undertaking service provision or statutory decision-making with direct effect on citizens

    b. establish, disseminate and regularly review the terms and conditions for the use of ADS and related systems

    c. develop easy-to-use templates, impact assessments, and conditions for public policy ADS.

    d. right of review and the ability to prevent or halt public policy ADS should any system be found in breach of the ethical standards they set out

    e. act as ombudsman for citizens affected by public policy ADS.

[40] Lepri, B. et al. 2017. "Fair transparent and accountable algorithmic decision-making processes" *Philosophy & Technology* http://www.nuriaoliver.com/papers/Philosophy_and_Technology_final.pdf

Annex I

Detailed considerations for ethical standards provision in public policy ADS

This list is indicative rather than exhaustive or prescriptive. It indicates that securing ethical standards in ADS is complex and difficult but not impossible. Likely considerations at a range of stages of ADS might include:

1. Decision-to-adopt ethical framework that sets out

   a. Minimum requirements for ethical standards to be met for different types of public policy ADS.
   b. Weighting to be given to concerns about objectivity and transparency according to the significance of using ADS in any one case.
   c. Highly staged public procurement processes that facilitate ethical review and exit points should the ethical requirements not be met.

2. Decision / inference model and source code guidelines that set out

   a. Requirements for publication of the originating model and assumptions; open-source code
   b. Any iterative adjustments made during training of the ADS, with a special emphasis on mathematical adaptations to reduce 'bias'
   c. Required external validation of the quality, compatability and appropriateness of originating decision-model, source code and/or learning system

3. Data use, compilation, sharing, cleaning and storage guidelines that set out[41]

   a. Required external validation of the quality, compatability and appropriateness of training data.
   b. Any ethically-based exclusions on data use[42]
   c. Terms for unique use of any data so that it is confined to the ADS operation[43]
   d. Requirements for transparent explanation of all training data[44]
   e. Requirements that each public policy ADS can be terminated, with all data produced during its operation intact and stored solely in a secure and accessible (non-proprietary) format by the public body following termination

4. Functioning, application and audit guidelines that set out

   a. The status of any ADS recommendation or inference about individuals.
   b. Procedural requirements for decision reviews and systems audits of ADS recommendations, their use by human decision-makers and decision outcomes for individuals and groups.
   c. Requirements for evaluation and audit of the ADS user-interface, and its implications for how human decision-makers interpret machine-learning inferences.

---

[41] These are related to, but not completely covered by data ethics (in the UK, the responsibility of the Centre for Data Ethics).
[42] These could include exclusions by **type**, e.g. numerical, biometric, language; **format**, e.g. facial recognition, genetic, speech; **status**, e.g where consent might apply; or **origin**, e.g. obtained by secondary government agency or other organisation.
[43] ie., the data cannot be used to train additional ADS systems by the public body concerned, a private provider or other actor
[44] Including but not confined to, its original sources; cleaning and harmonisation; detail of meta-data.

SUBMISSION 13

Committee on Standards in Public Life
Artificial Intelligence (AI) and Public Standards Review

Written Submissions on behalf of Robin Allen QC and Dee Masters

www.ai-lawhub.com
www.cloisters.com

Introduction

1. We are specialist discrimination barristers based in London at Cloisters chambers with deep knowledge of the equality implications of artificial intelligence (AI) and algorithmic discrimination having written and lectured extensively on this area and created the AI Law Hub which is available at www.ai-lawhub.com. We also have extensive experience working alongside governmental and non-governmental bodies concerned with law making within the UK and Europe. An account of our relevant professional experience is set out in Appendix 1.

2. We are aware that standards in public life can be affected by the use of artificial intelligence (AI) and machine learning (ML). There is a commonly understood risk that both AI and ML can lead to non-human-centric outcomes which would be wholly at odds with all the currently stated ethical standards and will, in undermining respect for human dignity, undermine confidence in public administration.

3. Many of these issues are discussed on our website at both https://ai-lawhub.com/business/ and also https://ai-lawhub.com/government/ We invite the Committee to review those sites.

4. The terms of reference for this review by the Committee propose it should "Examin[e] the current use of artificial intelligence and associated advanced technologies in the public sector." In the past 2-3 years, commentators within Europe and globally have slowly become aware of the potential for algorithmic discrimination. Unsurprisingly, many governments are taking tentative steps towards legislating in this area (https://ai-lawhub.com/government/) and private companies are voluntarily creating codes and charters in order to regulate their own activities (https://ai-lawhub.com/business/).

5. While there is an increasing awareness at the governmental level that "something must be done"; we note that the incoming European President of the European Commission Ursula von der Leyen has promised to propose legislation in this field within her first 100 days, on the other hand the degree to which this is recognised by business is more patchy and we believe that sometimes it is more defensive and self-interested, and based on the view that it is inevitable that regulation in this area will be developed.

6. So, the Review is greatly to be welcomed.

7. We have outlined many examples across the full spectrum of equality law in the public and private sector on our website at https://ai-lawhub.com/framing-the-debate/. Rather than set out a further long catalogue of the known occasions on which this has occurred we refer the committee to our twitter feed @AIlawhub where many examples will be found.

8. Considering our interest in algorithmic discrimination, we shall focus our remarks on this aspect. We set out below a summary account of algorithmic discrimination in the public and private sector and an analysis of the existing legal framework within the Equality Act 2010 (EA 2010) and the General Data Protection Regulation (GDPR) before outlining six areas where we consider that the Government could introduce measures which would help tackle discrimination arising from AI in the public sphere.

Existing legal framework

Equality Act 2010

9. Nothing in the EA 2010 specifically refers to AI or ML. This seems very odd now since it is well understood that AI decisions and ML can be discriminatory if either the data set is skewed or the algorithm operates in a biased way. However the Act was passed before the awareness of this kind of problem had come to the fore.

10. In fact there is no specific equality legislation in the UK or within European law, that applies specifically to AI or ML rather there are general provisions that can be seen to give some protection specifically against algorithmic discrimination in the provision of goods, facilities and services and when organisations carry out public functions.

11. This is a key point to be noted. While there is much discussion about the ethical framework within which AI and ML should work so far the discussions about the legislative regulatory framework to prohibit discrimination are very limited indeed. What can the Equality Act 2010 do?

12. The principal provisions of the EA 2010 follow a well-defined pattern -

a. First the Act defines certain human characteristics as being "protected characteristics": age; disability; gender reassignment; marriage and civil partnership; pregnancy and maternity; race; religion or belief; sex; sexual orientation.

b. Next it states that certain kind of behaviour is "prohibited discriminatory conduct", in relation to a protected characteristic.

c. Then it states the particular circumstances in which prohibited discriminatory conduct is "unlawful".

d. It then explains the way in which complaints can be made.

e. Lastly it sets out specific exceptions.[1]

13. We shall explain how this Act works as it might relate to AI and ML in a little greater detail -

a. The EA 2010 applies (subject to the limited exceptions) to all individuals and organisations that provide goods, facilities and services or exercise a public function by virtue of section 29. It also applies to range of activities elsewhere such as work and education.

b. EA 2010 section 13 defines the first class of prohibited discriminatory conduct as direct discrimination; this will occur when someone is treated less favourably because of a protected characteristic. Importantly, if a rule or provision is applied which means that everyone who is disadvantaged by it shares a particular protected characteristic, and everyone who is not disadvantaged by that rule or provision does not possess the protected characteristic, then direct discrimination will have occurred. A detailed exposition of these type of "proxy" direct discrimination claims is available here. Other than age discrimination, direct discrimination can never be justified and will always be unlawful unless an exception contained within the EA 2010 applies.

d. Indirect discrimination is defined as prohibited discriminatory conduct by EA 2010 section 19 where a person (A) applies to another person (B) a provision, criterion or practice which is applies or would apply to everyone, but it puts or would put persons with whom B shares a protected characteristics at a particular disadvantage when compared with persons with whom B does not share it, and B is at this disadvantage and A cannot show it to be a proportionate means of achieving a legitimate aim.

e. Harassment is also is defined as prohibited discriminatory conduct by EA 2010 section 26. It will occur when a person (A) engages in unwanted conduct related to a relevant protected characteristic, and the conduct has the purpose or effect of violating B's dignity, or creating an intimidating, hostile, degrading, humiliating or offensive environment for B.

f. EA 2010 sections 20 - 22 also impose obligations upon employers, service providers and public authorities to make reasonable adjustments. This means that where

[1] See the appendices to EA 2010. Those exceptions apply to certain specific and limited classes of activity; they are discussed shortly below.

a provision, criterion or practice of A's puts a disabled person at a substantial disadvantage in comparison with persons who are not disabled, A must take such steps as it is reasonable to have to take to avoid the disadvantage.

14. Whilst we are not aware of the EA 2010 being deployed to challenge discriminatory algorithms within the domestic court system (as opposed to the United States), there are many ways in which we consider the EA 2010 could be utilised. Our analysis is set out in detail at https://ai-lawhub.com/framing-the-debate/.

15. Finally, it is important to note that among the variety of exceptions contained within the EA 2010 which limit the scope of the principle of non-discrimination there are two of note for this paper: where there is an exercise of a judicial function (para 3 in Part 1 of Schedule 3 of the Equality Act 2010 which is here) and where age is taken into account in relation to the provisions of financial services (para 20a in Part 1 of Schedule 3 which is here).

General Data Protection Regulation

16. The key legislation in Europe expressly relating to algorithms, machine learning and data protection is the General Data Protection Regulation (GDPR) and the Legal Enforcement Directive. The GDPR is transposed into UK law via the Data Protection Act 2018(DPA 2018).

17. Our analysis of this legal framework and the extent to which individuals can object to automated decision making under Articles 21 and 22 is set out at https://ai-lawhub.com/data-protection-existing-legal-framework/ which contains an exploration of the extent to which the GDPR enshrines an effective principle of transparency.

Steps to further address algorithmic discrimination in the public sector

18. We consider that there are at least six steps that the Government could take in order to tackle algorithmic discrimination in the public sector. The first two need little argument here since they are widely recognised limitations on the controls in the GDPR.

A. **Extend GDPR Article 22 and the right to object to automatic profiling to all decisions which produce legal effects for data subjects or similarly significantly affects the data subject.**

19. If we are to get serious about data protection in the era of AI and ML. We need to think hard about proceeding in this direction.

B. **Extend GDPR Article 21 and the right to object to automatic profiling to circumstances beyond Article 6 (1)(e) and / or (f)**

20. Once again, we think that the protection of personal data is so important to avoid discrimination and abuse that there should be an extension of Article 21.

C. **Create a meaningful principle of transparency**

21. One of the key challenges for bodies seeking to regulate algorithms and machine learning technology is that the data which they use and the basis upon which decisions are made are likely to be hidden within "the black box" meaning that it is difficult to create meaningful accountability.

22. The GDPR enshrines a principle of transparency, which in theory, could be utilised to ensure that algorithms and the data that they use is readily understandable to the public. However, the current Information Commissioner's Office (ICO) guidance entitled "Automated decision-making and profiling" suggests that the principle of transparency is weak since it provides little by way of an obligation to provide meaningful information:

> **How can we explain complicated processes in a way that people will understand?**
>
> Providing 'meaningful information about the logic' and 'the significance and envisaged consequences' of a process doesn't mean you have to confuse people with over-complex explanations of algorithms. You should focus on describing:
>
> - the type of information you collect or use in creating the profile or making the automated decision;
> - why this information is relevant; and
> - what the likely impact is going to be/how it's likely to affect them.
>
> Example
> An on-line retailer uses automated processes to decide whether or not to offer credit terms for purchases. These processes use information about previous purchase history with the same retailer and information held by the credit reference agencies, to provide a credit score for an online buyer.
>
> The retailer explains that the buyer's past behaviour and account transaction history indicates the most appropriate payment mechanism for the individual and the retailer.
>
> Depending upon the score customers may be offered credit terms or have to pay upfront for their purchases.

23. Our view is that this interpretation of the GDPR is unlikely to compel organisations to be open and transparent about algorithms and the basis for machine learning.

24. In order to create accountability in public life, we consider that the ICO guidance entitled "Automated decision-making and profiling" should be revisited so as to create a more meaningful form of transparency whereby Data Subjects have the right to receive a detailed explanation of the way in which their data is being processed beyond a generic description.

D. **Removing the judicial functions exception within the EA 2010 in relation to algorithms**

25. In the US, algorithms are also being used in relation to sentencing decisions. The most famous example relates to an algorithm used within software called Compas. This is used in some states by judges to inform sentencing decisions. This has led commentators such as journalists working for Propublica to analyse whether the Compas software creates discriminatory outcomes. Propublica concluded that black defendants were twice as likely to be incorrectly labelled as high risk offenders by Compas. It is denied by Compas' makers that its technology is discriminatory.

26. Whilst we are not aware of this type of technology being used in the UK, it is important to note that it would probably not infringe the EA 2010 because of the exception pertaining to judicial functions in para 3 in Part 1 of Schedule 3 which reads as follows:

> (2) Section 29 does not apply to:
> (e) a judicial function;
> (f) anything done on behalf of, or on the instructions of, a person exercising a judicial function;
> (g) …
> (h) …
> (3) A reference in sub-paragraph (1) to a judicial function includes a reference to a judicial function conferred on a person other than a court or tribunal.

27. There is no definition of "judicial function" within the EA 2010 beyond this provision. However, there are some relates sources of information which suggest that the "judicial function" exception is intended to capture merits-based decisions reached by judges and persons in a similar position. In particular, the Explanatory Notes that accompany the EA 2010 explains that:

> A decision of a judge on the merits of a case would be within the exceptions in this Schedule. An administrative decision of court staff, about which contractor to use to carry out maintenance jobs or which supplier to use when ordering stationery would not be.

28. There is further guidance from the Equality and Human Rights Commission in its document entitled, "Your rights to equality from the criminal and civil justice systems and national security" where the distinction between a judicial function and related decisions is unpicked. The following passage is material:

> Equality law does not apply to what the law calls a judicial act. This means something a judge does as a judge in a court or in a tribunal case. It also includes something another person does who is acting like a judge, or something that they have been told to do by a judge.
>
> For example: A father, who is a disabled person who has a visual impairment, applies to court for a residence order in respect of his child. The court refuses his application. He believes that this is because of his impairment. As the decision of the court is a judicial act, he may be able to appeal against the decision, but he cannot bring a case against the judge under equality law.
>
> …

If the disabled person feels that he or she has been treated unfavourably in subsequent dealings with the Crown Prosecution Service or, in Scotland, the Procurator Fiscal's office, for example if they refuse to call him as a witness because they think he will not present well to the jury because of his learning disability, or if the CPS only offers to meet him a place which is inaccessible to him without making reasonable adjustments, then they may well be able to bring a claim for unlawful discrimination under equality law.

29. On this basis, technology like Compas could be utilised in the UK without falling foul of the EA 2010. Considering Propublica's research, this is an area which is likely to consider urgent consideration in the near future if algorithms start to be used in the UK's legal system in relation to judicial decisions like sentencing. The alternative is that the public may start to lose trust and faith in the judicial system.

E. **Removing the financial services exception within the EA 2010 in relation to age**

30. This is a source of constant complaint to our own knowledge. The exception enables the kind of "compute says no" refusal to grant insurance or other financial products to those who are of and above a certain age. The actual age can vary considerably from 65 upwards. The individual has no idea as to why the decision has been taken and frequently it cannot be explained by any customer facing staff.

31. Decisions such as these commonly are based on algorithms that have been designed to discriminate on the basis of past knowledge about health or mortality for instance. As it is well known that there is a broad range around any mean statement of capacity or mortality these can be highly discriminatory weeding out good risks with the bad in an indiscriminate way. They are not remotely human-centric and certainly undermine human dignity.

F. Developing a statutory Code of Practice on the use of AI in the public setting

32. There are numerous statutory Codes of Practice in existence which are intended to assist the public, organisations and the courts understand the current legal system. An example of a statutory Code of Practice in the public sector is a document produced by the Equality and Human Rights Commission entitled "Services, Public functions and Associations: Statutory Code of Practice" which is available here.

33. Since algorithmic discrimination is a relatively new concept, there is no equivalent Code in relation to the use of AI within the public sector. The most detailed guidance to date has been by the ICO (referred to above) but inevitably this analyses only one form of AI namely automated profiling and from a purely data protection perspective. In light of the potential for AI transform public life, we strongly recommend that a new statutory Code of Practice is formulated which explores permissible uses of AI in the public sector and provides detailed guidance for organisations carrying out public functions on how to conduct risk assessments and audits so as to evaluate the appropriateness of any proposed technology.

Conclusion

26. We commend the work of the Committee and would be happy to come and discuss any part of this paper or any issue or matter that may be helpful to the C0mmittee's Review

ROBIN ALLEN QC
DEE MASTERS

AI-LAWHUB.COM
CLOISTERS

2nd August 2019

Appendix
Relevant professional experience

1. Robin Allen QC: I am a discrimination barrister from Cloisters who has appeared in over 145 reported cases in the Industrial Relations Law Reports (more than any other barrister). These include many path-finding appellate cases at the highest level in the UK and Europe. I have been instructed in over 40 cases in the House of Lords/Supreme Court and have undertaken test cases in relation to every protected characteristic. I have worked with the European Commission on the drafting of the two key equality Directives in 2000, and lectured on their meaning to judges and NGOs and indeed Equinet members. A copy of my full CV is available here.

2. Dee Masters: I am also a discrimination barrister from Cloisters. My practice primarily consists of multi-week trials involving multiple and complex claims. Due to my expertise, I have advised NGOs and government agencies on the development of discrimination law and I have delivered judicial training on discrimination law at Academy of European Law (ERA) in Trier funded by the European Commission. A copy of my full CV is available here.

3. Together, we started the debate on the ways in which the Equality Act 2010, which enshrines the principle of non-discrimination in the UK, could be deployed to challenge discriminatory technology and in particular algorithms. We have published several papers and spoke at numerous events, some of which are highlighted below:

   a. Robin Allen QC and Dee Masters: "Algorithms, Apps & Artificial Intelligence: The next frontier in discrimination law?", October 2018, available here.

   b. Robin Allen QC and Dee Masters: "Algorithms, apps & artificial intelligence 2: Can data protection laws be used to challenge discriminatory tech?", July 2018, available here.

   c. Robin Allen QC: ER*A*, Trier, June 2018, paper available here.

   d. Public Law Project, "AI Justice: Artificial intelligence decision-making and the law", October 2018, London, details available here.

   e. Robin Allen QC, April 2019, ERA, Brussels, details available here.

   f. Robin Allen QC and Dee Masters: "Algorithms, Apps and Artificial Intelligence", Discrimination Law Association, July 2018, details available here.

4. In 2019, we created www.ai-lawhub.com which is a centralised resource for tracking and analysing the emerging debate in the UK, Europe and globally concerning AI and the ways in which it should be analysed with an equality, data protection and human rights framework. We also tweet regularly from @AILawHub on developments in this area.

5. Beyond our work in the field of AI and discrimination, we have both worked (separately and together) for governmental and non-governmental bodies concerned with law making in the past as follows:

   a. Robin Allen QC and Dee Masters: Advised the Equality Commission of Northern Ireland (ECNI) and the Northern Ireland Commissioner for Children and Young People (NICCY) in April 2013 on its response to proposed legislation extending the existing prohibition on age discrimination to children and young people. The opinion was published and is available here.

   b. Robin Allen QC and Dee Masters: Advised AGE Platform in October 2014 on a proposed European Directive which will, if enacted, extend the prohibition on age discrimination into the field of goods, facilities and services. AGE Platform is an NGO which promotes the interests of people over 50 in Europe. More information is available here.

   c. Dee Masters: Keynote speaker at a conference August 2015, by invitation from the ECNI, attended by politicians and policy makers in Belfast concerning extending age discrimination legislation so as to protect children and young people. A copy of the paper is here.

   d. Robin Allen QC and Dee Masters (and other colleagues at Cloisters): Drafted the Technical Guidance accompanying the ban on age discrimination in goods, facilities and services for the EHRC in GB. The final version published by the EHRC in March 2016 is available here.

   e. Robin Allen QC and Dee Masters: Submitted evidence to the Joint Committee on Human Rights in September 2016, along with Robin Allen QC, in relation to age discrimination and children. A copy of the submission is available here.

   f. Robin Allen QC and Dee Masters: Advised AGE Platform in 2017 on the content of a proposed new UN Convention on age discrimination along with Robin Allen QC.

   g. Robin Allen QC and Dee Masters: Co-authors of a report to the UN's Open-ended Working Group on Ageing: Equality and non-discrimination along with academics and policy makers who specialise in age discrimination in the UK and internationally. Paper available here.

h. Robin Allen QC and Dee Masters: Participated in an internal meeting chaired by the UN Team within the Foreign & Commonwealth Office on 17 May 2017 in which the UN's call for evidence in relation to age discrimination was debated with academics, charities and other parties interested in combating age discrimination.

i. Robin Allen QC and Dee Masters: Chair and rapporteur respectively on 7 June 2017 at AGE Platform's annual conference in Brussels entitled, "Inequalities and abuse in old age: Time to Act!" for the session called, "Equality and non-discrimination in old age". Programme available here.

j. Dee Masters: Gave expert evidence to the UK Parliament's Women and Equalities Committee on older people in the workplace in 2018 which was subsequently featured in its report, available here.

k. Robin Allen QC and Dee Masters: Chair and speaker respectively on 25 January 2019 at an event hosted by Age UK and AGE International entitled, "Why the time is now right for a Convention on the rights of older persons".

l. We have lectured on issues relating to equality and non-discrimination on many occasions across Europe including to the Academy of European Law at Trier.

m. Robin Allen QC has co-authored Equality Law and Human Rights (3rd Edition 2018, Oxford University Press), and later this year Cambridge University Press will publish the developed texts of his Hamlyn Lectures provisionally entitled "Making Comparisons in Equality Law: Within gender, age, and conflicts"

SUBMISSION 14

CarnegieUK TRUST

CHANGING MINDS • CHANGING LIVES

RESPONSE TO COMMITTEE ON STANDARDS IN PUBLIC LIFE REVIEW OF ARTIFICIAL INTELLIGENCE AND PUBLIC STANDARDS

1. This short note makes a primary factual observation and covers a full reference paper that sets out work we have carried out to develop a proposal for a statutory duty of care for harm reduction on social media. We would commend this work to the Committee as they start their deliberations for this important review.

<u>AI safeguards and the accountability principle of public life – where are the Health and Safety at Work Act 1974 risk assessments?</u>

2. William Perrin, Trustee of Carnegie UK Trust. worked with Lord Stevenson of Balmacara to establish whether the provision of the Health and Safety at Work Act (HSAWA) 1974 applied to artificial intelligence systems used in decision support that may be 'articles supplied for use at work' in Section 6 of the Act. The Government is adamant that it does. And that this 'requires such testing and examination as may be necessary to ensure that any article for use at work is safe and without risks'. The exchange with Baroness Buscombe is below. In our work on online harms we have found few who understand that this existing regulatory regime applies to AI/ML etc.

3. The Committee might wish to follow up on the spirit of the accountability principle by asking government bodies for risk assessments and testing they have had carried out on AI systems and discuss the matter with the Health and Safety Executive. More broadly the Committee might wish to consider how the statutory duties of care in the HSAWA 1974 are being implemented by the government when they deploy AI and ML decision support systems.

   Lord Stevenson of Balmacara 23 May 2018 to Department for Work and Pensions[1]

   Industrial Health and Safety: Artificial Intelligence HL8200

   To ask Her Majesty's Government what assessment they have made of the extent to which section 6 of the Health and Safety at Work etc. Act 1974 applies to artificial intelligence or machine learning software that is used in the workplace to (1) control or animate physical things in the workplace, (2) design articles for use in the workplace, or (3) support human decision-making processes running on computers under the control of the employer with an impact on people's health and safety; and whether, in each case, testing regimes exist as set out in section 6(1)(b) of that Act.

   Answered by: Baroness Buscombe 05 June 2018

[1] https://www.parliament.uk/business/publications/written-questions-answers-statements/written-question/Lords/2018-05-23/HL8200/

Section 6 of the Health and safety at Work etc. Act 1974 places duties on any person who designs, manufacturers, imports or supplies any article for use at work to ensure that it will be safe and without risks to health, which applies to artificial intelligence and machine learning software. Section 6(1)(b) requires such testing and examination as may be necessary to ensure that any article for use at work is safe and without risks but does not specify specific testing regimes. It is for the designer, manufacturer, importer or supplier to develop tests that are sufficient to demonstrate that their product is safe.

The Health and Safety Executive's (HSE) Foresight Centre monitors developments in artificial intelligence to identify potential health and safety implications for the workplace over the next decade. The Centre reports that there are likely to be increasing numbers of automated systems in the workplace, including robots and artificial intelligence. HSE will continue to monitor the technology as it develops and will respond appropriately on the basis of risk.

A statutory duty of care for social media harm reduction

4. In 2018-2019, Professor Lorna Woods (Professor of Internet Law in the School of Law at the University of Essex) and William Perrin (a Carnegie UK Trustee and former UK government Civil Servant) developed a public policy proposal to improve the safety of some users of internet services in the United Kingdom through a statutory duty of care enforced by a regulator. Woods and Perrin's work under the aegis of Carnegie UK Trust took the form of many blog posts, presentations and seminars.

5. A full reference paper drawing together their work on a statutory duty of care was published in April 2019, just prior to the publication of the Online Harms White Paper. We have attached it as an annex to this paper and it can be viewed, along with all the other material relating to this proposal and a full recent response to the DCMS consultation on the Online Harms White Paper, on the Carnegie UK Trust website: https://www.carnegieuktrust.org.uk/project/harm-reduction-in-social-media/

6. Our work has influenced the recommendations of a number of bodies, including: the House of Commons Science and Technology Committee, the Lords Communications Committee, the NSPCC, the Children's Commissioner, the UK Chief Medical Officers, the APPG on Social Media and Young People and the Labour Party.[2] A statutory duty of care has been adopted – though not fully as we envisaged – by the Government as the basis for its Online Harms White Paper proposals[3]. Most recently, though it did not refer to our work, a report to the French Ministry of Digital Affairs referenced a "duty of care" as the proposed basis for social media regulation.[4]

---

[2] https://www.nspcc.org.uk/globalassets/ documents/news/taming-the-wild-west-web-regulate-social-networks.pdf; https://www.childrenscommissioner.gov.uk/2019/02/06/childrens-commissioner-publishes-astatutory-duty-of-care-for-online-service-providers/; https://www.gov. uk/government/publications/uk-cmo-commentary-on-screen-time-and-social-media-map-ofreviews/; https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/822/82202.htm; https://labour.org.uk/press/tom-watson-speech-fixing-distorted-digital-market/; https://www.parliament.uk/business/committees/committees-a-z/lords-select/ communications-committee/inquiries/parliament-2017/the-internet-to-regulate-or-not-toregulate/; https://www.rsph.org.uk/our-work/policy/wellbeing/new-filters.html

[3] https://www.gov.uk/government/consultations/online-harms-white-paper

[4] http://www.iicom.org/images/iic/themes/news/Reports/French-social-media-framework---May-2019.pdf

7. We urge the Committee and its review team to read our reference paper in full. Any discussion of how to set standards for AI use in any sector will need to take account of a wide range of technical and ethical issues. But we believe that taking a step back may help consider the issues in a simpler way: our proposition for a systemic duty of care to reduce reasonably foreseeable harms occurring on social media – based on well-established frameworks in areas such as health and safety – is a good starting point for all considerations of how to ensure the safe, ethical and fair deployment of any technology that has an impact on end users.

8. There are three particular areas where we believe our work has relevance for any considerations around the deployment of AI and the impact on public life:

    i) AI as the algorithm that promotes content. From a duty of care perspective, we would see this as driving the sorts of content that is promoted and the sorts of content that is excluded from searches/autoplay, all of which are design choices made by the platform or service provider, and within this we would also consider tools that are available from third parties eg that allow a user to search hashtags;
    ii) the role of AI in creating content – eg the technology behind deepfakes; while this is not directly covered in our work, it is flagged in the DCMS Online Harms White Paper as a contributing factor to harms caused by disinformation, by which it is becoming even easier to create and disseminate false content and narratives (p23)
    iii) the role of AI in spotting the problem, eg identifying and removing illegal or harmful content, where we see a number of particular concerns:

        - the focus on this makes it an *ex post* issue when we would see a duty of care requiring companies to consider more basic questions of platform design;
        - the material that is being used for training: there are possible problems with an unequal coverage (eg language but also bias);
        - does the focus on AI distract from other questions (eg focus on design choices)?

9. There is also an intersection between the Committee's inquiry and the Online Harms White Paper in relation to emergence of harms to democracy. Along with other civic society organisations, we have published a statement calling for a greater focus on societal harms, such as those which impact democracy, in the Government's White Paper proposals and the scope of harms to come under the regulatory regime.

10. Our work draws on a number of established legal and policymaking frameworks but we would particularly draw the Committee's attention to our discussion, in chapter two of the attached paper, on the application of the precautionary principle to the question of how to regulate or set standards for innovative, fast-developing technologies:

    The government has often been called to act robustly on possible threats to public health before scientific certainty has been reached. After the many public health and science controversies of the 1990s, the UK government's Interdepartmental Liaison Group on Risk Assessment (ILGRA) published a fully worked-up version of the precautionary principle for UK decision makers: 'The precautionary principle should be applied when, on the basis of the best scientific advice available in the time-frame for decision-making: there is good reason to believe that harmful effects may occur to human, animal or plant health, or to the environment; and the level of scientific uncertainty about the consequences or

likelihoods is such that risk cannot be assessed with sufficient confidence to inform decision-making.' [5]

The ILGRA document advises regulators on how to act when early evidence of harm to the public is apparent, but before unequivocal scientific advice has had time to emerge, with a particular focus on novel harms. ILGRA's work focuses on allowing economic activity that might be harmful to proceed 'at risk', rather than a more simplistic, but often short-term politically attractive approach of prohibition. The ILGRA's work is still current and hosted by the Health and Safety Executive (HSE), underpinning risk-based regulation of the sort we propose.

11. We would also particularly refer the Committee to our discussion, in chapter 3, on the design decisions that drive any of our interactions with technology – whether viewing content online, or engaging with AI-enabled services. These environments are defined by code that the service providers have actively chosen to deploy, their terms of service or contract with the user and the resources service providers deploy to enforce that. While technological tools can be used for positive reasons as well as have negative impacts, it is important to remember that they are not neutral,[6] nor are they immutable. Corporate decisions drive what content is displayed to a user. Service providers could choose not to deploy risky services without safeguards or they could develop effective tools to influence risk of harm if they choose to deploy them.

12. These decisions are best taken when informed by a risk assessment. There will be risks which will be obvious – for instance material harm is known to have occurred before in certain circumstances and those which, while not obvious are foreseeable. If a material risk is foreseeable then a company should take reasonable steps to prevent it. This is as true – and even more important – in relation to the deployment of AI as it is in relation to any other technology or service.

13. There are many moving parts in this landscape, and many government and regulatory organisations undertaking concurrent reviews of bits of it. Protecting users from harm – however it manifests itself - has to be at the heart of all those proposals. Given the commitment of the Government to introduce a statutory duty of care to reduce online harms, we would urge the Committee on Standards in Public Life to consider how their review might take account of its principles and consider how the thinking that underpinned our work can apply to theirs.

14. We are happy to speak to you further about our proposals or assist in any way in the Committee's review.

Carnegie UK Trust

July 2019

[Attachment: "Online Harm Reduction: a statutory duty of care and a regulator" (April 2019)]

---

[5] United Kingdom Interdepartmental Liaison Group on Risk Assessment (UK-ILGRA), The Precautionary Principle: Policy and Application, available: http://www.hse.gov.uk/aboutus/meetings/committees/ilgra/pppa.htm

[6] W. Hartzog, Privacy's Blueprint: The Battle to Control the Design of New Technologies (Cambridge, MA: Harvard University Press, 2018)

SUBMISSION 15

Comments by the Centre for Information Policy Leadership on the UK's Committee on Standards in Public Life Review of Artificial Intelligence and Public Standards

The Centre for Information Policy Leadership (CIPL) is an independent global data privacy and cybersecurity think tank with over 70 member companies. Its mission is to engage in thought leadership and develop best practices that ensure both effective privacy protections and the responsible use of personal information in the modern information age. CIPL's work facilitates constructive engagement between business leaders, privacy and security professionals, regulators and policymakers around the world.[24]

CIPL welcomes the opportunity to contribute to the Review of Artificial Intelligence and Public Standards which has been launched by the UK's Committee on Standards in Public Life.[25]

Although a great deal has been written about Artificial Intelligence in recent years, CIPL believes that this Review is the first to focus on the ethical issues which are, and increasingly will be, raised by the use of AI specifically by public authorities and those delivering public services. CIPL congratulates the Committee for this important initiative.

This submission does not aim to be comprehensive. Instead, we draw attention to some of the relevant work that CIPL has already undertaken. We hope that some of the information and analysis we have assembled may be of assistance to the Committee, particularly our observations about the relationship between AI and data protection regulation. We also suggest that there may be some ideas, approaches and techniques which could usefully be adopted and adapted by the Committee for its Review.

Delivering Sustainable AI Accountability in Practice

CIPL has been conducting extensive research on the interplay between AI and data protection through its project on "Delivering Sustainable AI Accountability in Practice".[26] This on-going project aims to provide a detailed understanding of the opportunities presented by AI, its challenges to data protection laws and practical ways to address these issues through best practices and organisational accountability. CIPL published its first White Paper "Artificial Intelligence and Data Protection in Tension"[27] in October 2018.

---

[24] CIPL is a global data privacy and cybersecurity think tank in the law firm of Hunton Andrews Kurth and is financially supported by the law firm and 77 member companies that are leaders in key sectors of the global economy. CIPL's mission is to engage in thought leadership and develop best practices that ensure both effective privacy protections and the responsible use of personal information in the modern information age. CIPL's work facilitates constructive engagement between business leaders, privacy and security professionals, regulators and policymakers around the world. For more information, please see CIPL's website at http://www.informationpolicycentre.com/. Nothing in this paper should be construed as representing the views of any individual CIPL member company or of the law firm of Hunton Andrews Kurth.

[25] Review of AI and Public Standards, UK Committee on Standards in Public Life, available at https://www.gov.uk/government/collections/ai-and-public-standards.

[26] See CIPL Project on Artificial Intelligence and Data Protection: Delivering Sustainable AI Accountability in Practice: https://www.informationpolicycentre.com/ai-project.html.

[27] See CIPL white paper on "Delivering Sustainable AI Accountability in Practice: Artificial Intelligence and Data Protection in Tension", 10 October 2018, available at https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_ai_first_report_-_artificial_intelligence_and_data_protection_in_te....pdf.

This paper includes:

- an Introduction to Artificial Intelligence, noting that the term encompasses a wide variety of current and prospective technological innovations, each of which presents distinct challenges to existing norms, policies and laws;
- a summary of the capabilities of Artificial Intelligence, looking particularly at the functions, potential and implications of machine learning, deep learning and Natural Language Processing (NLP);
- an overview of public and private uses of Artificial Intelligence, noting some of the changes in major sectors influenced by emerging AI technologies. These include healthcare, transportation, financial services, marketing, agriculture, education, cybersecurity, law enforcement and public services;
- an analysis of some of the tensions that AI brings – or will very soon bring – for data protection law and compliance.

This tensions identified in this latter section of the CIPL paper include:

- problems with definitions of "personal data";
- potential conflicts with limitations on collection, purpose specification and use of personal data;
- problems with excessively narrow approaches to data minimisation and data retention;
- challenges of transparency and openness, especially where the "black box" phenomenon calls into question how to provide notice about "the unpredictable and the unexplainable";
- the need for data quality, comprehensiveness and correction mechanisms;
- the particular challenges of laws which specifically address profiling and automated decision-making.

The paper concludes by elaborating six key observations, all of which have relevance to the Committee's Review:

- not all AI is the same;[28]
- AI is already widely used in society and is of significant economic and societal value;
- AI requires substantial amounts of data to perform optimally;
- AI requires data, including sensitive personal data, to identify and guard against bias;
- the role of human oversight of AI will need some re-definition for AI to deliver the greatest benefits;
- AI challenges some of the basic requirements of data protection law.

Ethical Implications of Artificial Intelligence

The CIPL White Paper notes repeatedly that the development and deployment of various AI techniques in various contexts increasingly raises ethical issues which go much wider than the focus of data protection regulation. There are also some very difficult tensions between benefits and risks. These concerns tie in with emerging debates about data ethics.

[28] A point forcibly made by the recent ICO/Turing Institute's "Project Explain" paper which emphasises the importance of context in explaining AI decisions.

For example, the 40th International Conference of Data Protection and Privacy Commissioners in October 2018 adopted a Declaration on Ethics and Data Protection in Artificial Intelligence.[29] CIPL very much welcomed this focus of attention and much of the substance of the Declaration. But we had reservations about some of the detailed points where we feared excessive threats to beneficial innovation. Accordingly CIPL published a Response to the Declaration[30] in January 2019. The Response argued the need for novel, flexible, risk-based and creative approaches to addressing the challenges, even if this means some departures from conventional interpretations of privacy principles. Moreover, while respect for privacy rights must be a key consideration in the development of AI, such rights are not absolute and must be balanced against other human rights, such as those respecting life and health, and the benefits of the AI to individual users and society as a whole.

Again, the Committee may find some parts of our detailed Response to be useful for its current Review.

Impact of Data Protection

The scope of the Committee's Review is wider than data protection, addressing some applications which will never involve any personal data at all. It is also narrower, focusing on the use of AI by public services alone. But the messages from the CIPL White Paper suggest that many innovatory and potentially beneficial uses of AI – perhaps the majority – will be severely threatened unless creative and flexible approaches are adopted towards data protection requirements. Such an approach needs to be adopted by regulatory bodies, but also by those (such as DPOs) overseeing compliance inside organisations. This is not to suggest that a creative, flexible approach means any sort of automatic green light. But nor should there be an automatic red light just because some sort of AI application is involved.

What will be important is to address these issues openly from the outset and establish true accountability of organisations when using AI: (1) having credible mechanisms to assess and demonstrate whether benefits outweigh downsides and (2) implementing an effective control system for each application which addresses both compliance and ethical considerations. As a minimum, regulatory bodies will expect such arrangements from accountable organisations if they are to be expected to adopt any sort of creative and flexible approach.

The remainder of this submission includes suggestions which the Committee could adopt to move towards such a situation.

Lessons from Data Protection

CIPL believes that there are many parallels and that it may in fact be possible to draw some lessons from successes with data protection as well as from innovations which have sought to address some of the limitations of a "traditional" approach.

---

[29] ICDPPC Declaration on Ethics and Data Protection in Artificial Intelligence, 23 October 2018, available at https://icdppc.org/wp-content/uploads/2018/10/20180922_ICDPPC-40th_AI-Declaration_ADOPTED.pdf

[30] See comments by the Centre for Information Policy Leadership on the International Conference of Data Protection and Privacy Commissioners Declaration on Ethics and Data Protection in Artificial Intelligence, 25 January 2019, available at https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_response_to_icdppc_declaration_on_ethics_and_data_protection_in_artificial_intelligence.pdf.

- A Principles-based approach: Although, as noted above, AI is increasingly calling into question the substance of some of the core data protection principles, CIPL believes that the basic concept of a principles-based approach is sound and has largely survived the test of time. Given such rapid changes in technology and the focus on public services, CIPL believes that it would be sensible for the Committee to articulate principles which public sector and third parties providers would be expected to observe when developing and deploying AI applications. This would avoid some of the drawbacks associated with excessively prescriptive legislative requirements, anticipate that public bodies can largely be expected to observe such principles without the need for legal enforcement and will enable rapid refinement in the light of experience.

  At this stage, CIPL hesitates to suggest a comprehensive set of such principles, although – as noted below – some of them will follow from the Standards Challenges which the Committee has already articulated. We do suggest, however, that it would be desirable and efficacious to set out some principles in explicitly negative terms – i.e. unacceptable processes and outcomes which the Committee believes that public service providers should avoid when using Artificial Intelligence.

- Accountability – Accountability is one of the Seven Principles of Public Life and, on any analysis, it will inevitably feature as a key element for ensuring clarity about how any AI application is being used. For over 10 years, CIPL has been the leading advocate of incorporating the Accountability Principle as a central component of any regulatory framework for data protection. It has now been incorporated into GDPR and into other data protection laws around the world, although its full implications have yet to be fully tested.

  The Accountability Principle requires organisations a) to implement a comprehensive privacy management program, that operationalises legal requirements into measurable rules and controls, b) be able to verify and continuously improve on implementation of such program and c) be able to demonstrate to regulator, oversight body, or individuals or general public the effectiveness of the program.

  The Committee may find it helpful to draw upon two papers on the subject which CIPL published in July 2018:

  - The Case for Accountability:[31] How it Enables Effective Data Protection and Trust in the Digital Society

  - Incentivising Accountability:[32] How Data Protection Authorities and Law Makers Can Encourage Accountability

---

[31] "The Case for Accountability: How it Enables Effective Data Protection and Trust in the Digital Society," 23 July 2018, available at https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_accountability_paper_1_-_the_case_for_accountability_-_how_it_enables_effective_data_protection_and_trust_in_the_digital_society.pdf.

[32] "Incentivising Accountability: How Data Protection Authorities and Law Makers Can Encourage Accountability", 23 July 2018, available at https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_accountability_paper_2_-_incentivising_accountability_-_how_data_protection_authorities_and_law_makers_can_encourage_accountability.pdf.

The two papers are summarised in a short introductory paper.[33] **In the papers, CIPL strongly advocates that both private sector and public sector organisations should be required to implement and demonstrate accountability. The following diagram was used to** show the essential elements of accountability which organisations must implement and continuously deliver and improve:

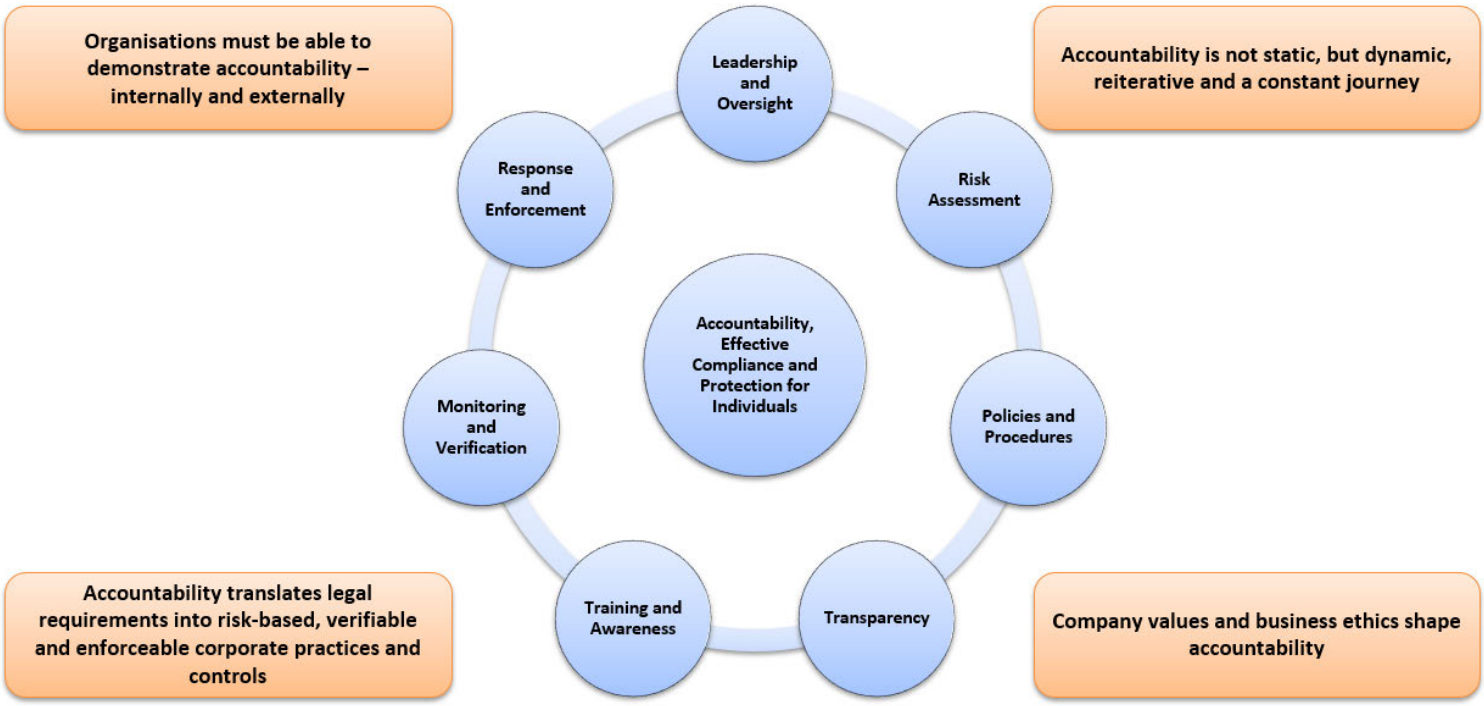


Figure I: CIPL Accountability Framework

- Governance, Leadership and Oversight – It is now widely recognised that data protection is a cultural challenge which cuts across many fragmented parts of an organisation. Good governance is essential to "get it right" so that leadership and oversight is needed at – or very near to – the top. The same must be true for acceptable use of Artificial Intelligence. Problems are inevitable if those at the top of public bodies do not have a good idea of which AI techniques are being used, what they are doing and how they are being governed and monitored.

  For many central UK public bodies, the Accounting Officer may need to certify annually that effective controls are in place to ensure that the AI Principles are being met. Similar controls may be needed for other bodies, such as NHS Trusts, police forces and local authorities.

  There may also be a need for external oversight. In the absence of a statutory regulator, there may be a role for the National Audit Office and other audit mechanisms.

- A risk-based approach – Since 2014, CIPL has argued the case for building into data protection laws an approach that puts particular emphasis on situations where the risks are

[33] Introduction to CIPL papers on the Central Role of Organizational Accountability in Data Protection", 23 July 2018, available at https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/introduction_to_two_new_cipl_papers_on_the_central_role_of_organisational_accountability_in_data_protection.pdf.

most likely or most serious. GDPR contains provisions which reflect this thinking, with particular reference to "risky" types of processing. CIPL's most recent paper on the subject - Risk, High Risk, Risk Assessments and Data Protection Impact Assessments[34] – unpacks the topic in more detail.

Any risk-based approach needs clarity about the threats and harms which could materialise and which need to be mitigated. The CIPL paper suggests how organisations should assess the likelihood and severity of any harms that might result from risky processing. The Committee may find it especially helpful to look at CIPL's thinking (at page 26) on the different types of harm. It is likely that similar harms will be identified as arising from unacceptable AI. The harms we classified were:

a) Material, tangible, physical or economic harm to individuals, such as:

- bodily harm;
- loss of liberty or freedom of movement;
- damage to earning power and financial loss; and
- other significant damage to economic interests, for example arising from identity theft.

b) Non-material, intangible distress to individuals, such as:

- detriment arising from monitoring or exposure of identity, characteristics, activity, associations or opinions;
- chilling effect on freedom of speech, association, etc.;
- reputational harm;
- personal, family, workplace or social fear, embarrassment, apprehension or anxiety;
- unacceptable intrusion into private life;
- unlawful discrimination or stigmatisation;
- loss of autonomy;
- inappropriate curtailing of personal choice;
- identity theft; and
- deprivation of control over personal data.

c) Societal harm – affecting Society in ways which go above and beyond individual harms, for example:

- damage to democratic institution;
- excessive state or police power; and
- loss of social trust ("who knows what about whom?").

- Impact Assessments – Whether or not a full risk-based approach is adopted, the Committee may wish to explore the benefits of Privacy and Data Protection Impact Assessments. These have been developed by various Data Protection Authorities (with the ICO playing a leading role) to encourage, and in some cases require, organisations to identify and document exactly how proposed data processing will impact individuals' lives. The GDPR now requires

[34] "Risk, High Risk, Risk Assessments and Data Protection Impact Assessments under the GDPR", 21 December 2016, available at https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_gdpr_project_risk_white_paper_21_december_2016.pdf.

the use of DPIAs in certain risky situations. CIPL finds that some of the leading private sector organisations have started leveraging successfully the DPIA methodology to conduct a wider human rights impact assessment or even an AI impact assessment.

The Committee could, for example, outline how the Impact Assessment concept could be modified and then used by public authorities before they use Artificial Intelligence in defined circumstances.

- Accountable AI practices – As a second stage of its Accountable AI Project, CIPL has embarked on finding emerging best practices that private sector organisations have started to build and deploy when developing AI technologies and applications. We have started to map some of those against CIPL's accountability framework to show that the same architecture of accountability may prove useful for those specific, additional AI controls and tools.

  We include the work so far on mapping of these nascent best practices, in case it proves useful to the Committee when coming up with examples of possible best practices in the public sector, too.

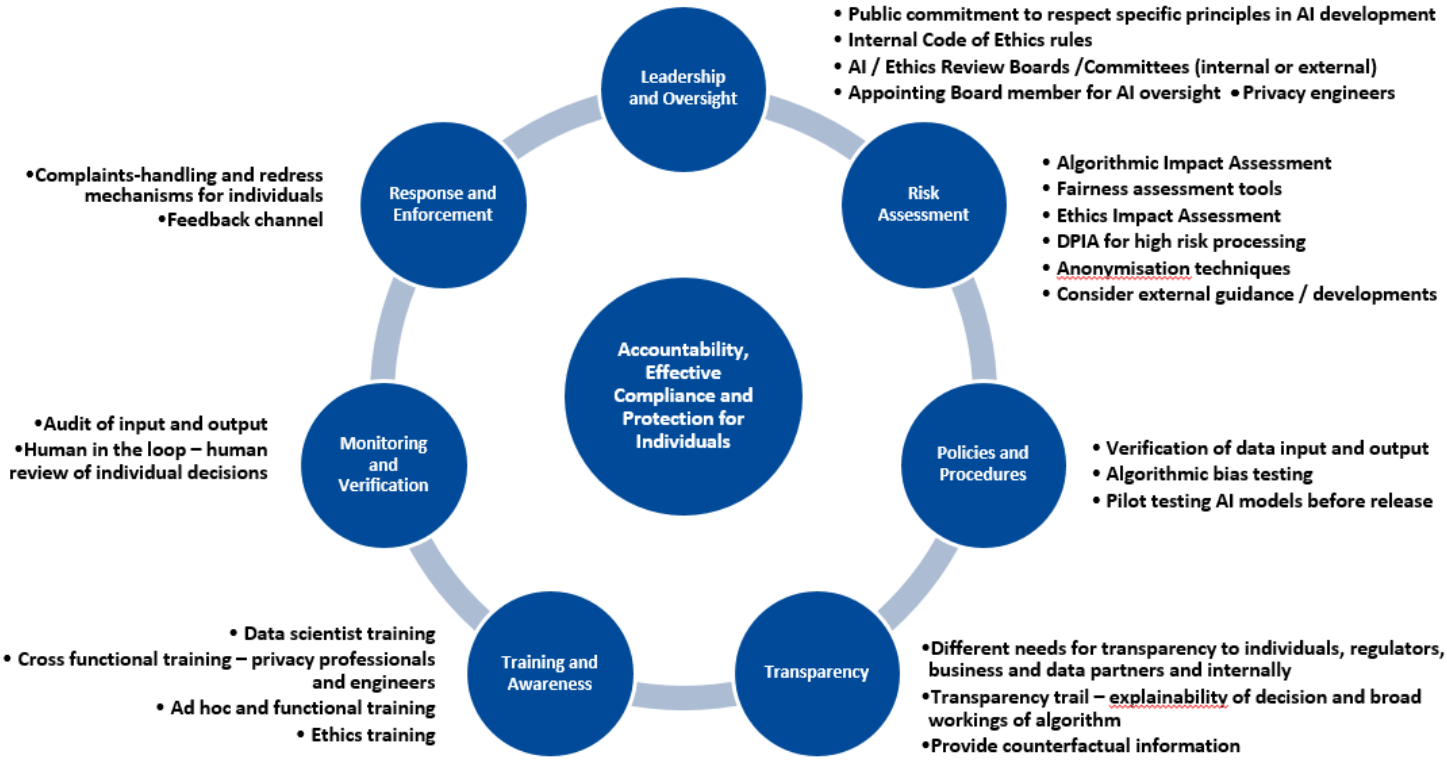


Figure II: AI Best Practices Mapped to CIPL's Accountability Framework

The Standards Challenges

CIPL is grateful to the Committee for sight of its paper which provisionally identifies six challenges which AI poses for public sector bodies.

These six challenges are broadly consistent with issues which have been discussed at length in various contexts by CIPL. The challenges need to be elaborated, but we are confident they will provide a solid foundation for a way forward.

We do, however, wish to make at this stage four comments on the paper:

- Context – The Committee may need to say more about features of public services which makes it especially important to establish a control framework to govern AI use. These include the monopolistic nature of many such services with no competitors and little need to safeguard reputation, the widespread use mandatory powers to collect and use information and the far-reaching nature of many public sector decisions which affect freedoms, rights and autonomies.
- Challenge 2 – The need to avoid bias goes further than demographic bias. Bias and discriminatory inferences can also arise from many other data sources – e.g. financial, health and consumption data and information about political and other opinions and behaviours.
- Challenge 5 – The risks of abuse are not limited to malicious acts. They can also come from ignorant, incompetent and sloppy administration on the part of officials and from inadequate resourcing.
- Extra Challenge – The paper is silent on the risks to democratic norms. Where AI means that decisions are made without adequate understanding about their rationale or where responsibility lies, there could be real threats to the principle of informed choice which must lie at the heart of the democratic process.

Principles of Good Decision-Making

As the Committee has made clear, AI will fundamentally change the way public bodies operate and deliver services. The impact on individual citizens from decisions which are made or assisted by AI use will be of particular concern. Although there will undoubtedly be many benefits, there will also be scope for unjustified or unfair processes and outcomes which could have dire effects.

The quality of decision-making by public bodies as it affects individuals tends to be a topic which is only debated when individual scandals surface. There is little systemic review. There is, however, disturbing evidence of existing failures to achieve "Right First Time", not least in the very high success rates of those who appeal to tribunals or other review mechanisms on such matters as social security, immigration and school exclusions. It is to be hoped that AI will improve this situation with greater "accuracy". But there are also risks that increasing reliance on AI may eliminate or marginalise human discretions, judgements and flexibilities. There are already frequent calls for more user-friendly and more humane public bureaucracies. It would be unfortunate if the quest for rationality reduced the human input further still.

Although adopted nearly 10 years ago before AI was on the agenda, CIPL draws the Committee's attention to the seven Principles which the Administrative Justice and Tribunal Council put forward for decision-making public bodies:

- Make users and their needs central, treating them with fairness and respect at all times;
- Enable people to challenge decisions and seek redress using procedures that are independent, open and appropriate for the matter involved;
- Keep people fully informed and empower them to resolve their problems as quickly and comprehensively as possible;
- Lead to well-reasoned, lawful and timely outcomes;

- Be coherent and consistent;
- Work proportionately and efficiently;
- Adopt the highest standards of behaviour, seek to learn from experience and continuously improve.

These Principles, which are elaborated in AJTC's 2010 report, drew upon the experience of the Parliamentary & Health Services Ombudsman in addressing maladministration by public bodies.

The Principles may need to be revived and made to resonate even more strongly as Artificial Intelligence is more and more used by public bodies which every day are making tens of thousands of decisions of real importance to citizens and their families.

SUBMISSION 16

CIPR CHARTERED INSTITUTE OF PUBLIC RELATIONS

The Committee on Standards in Public Life - AI and Public Standards Review
CIPR AI in PR Panel response
July 2019

For the public to adopt and realise the benefits of ethical artificial intelligence, it is imperative the UK Government devises a set of meaningful high level Public Standards to provide oversight and act as a governance safety check for the development, commissioning and deployment of the new technologies in Public Sector organisations, agencies and bodies.

This is vital at a time when public trust in institutions, particularly public institutions, is low. The CIPR AIinPR panel's role is to guide the global industry of public relations and communications specialists on how to best place and promote artificial intelligence as well as advise the government and its many public agencies and bodies on best practice for use of data and AI deployment.

The CIPR AIinPR panel believes for the public to understand and accept and adopt AI-enabled services the UK Government's AI and Public Standards for Artificial Intelligence must detail the following:

- Trust
- Transparency
- Ethics
- Fairness
- Accountability
- Leadership
- Safety checks

This must include effective engagement and dialogue with the public on the new ways AI-powered services will be provided and how data will be stored and used ethically and effectively. To build trust the UK Government's data processes for artificial intelligence-led services and tools must be transparent at all times, as does providing public body insight to improving value to the tax payer.

It is vital the use of individual's data complies with laws, such as GDPR, and the public have a clear understanding how their data may be used in any AI-enabled services, tools or to provide public services with insight to improve them, creating better value to the tax payer.

These include issues related to objectivity, transparency and accountability, as well as those around the appropriateness of AI-assisted decision-making in certain contexts, which may not be in the public interest, for example.

The CIPR AIinPR panel calls for the UK Government Standards on Artificial Intelligence to incorporate the requirement for the public to be clearly informed when they are interacting with an AI-powered service or tool, when they can interact with human assistants as well as when AI is used

85 Tottenham Court Road, London W1T 4TQ
T +44(0)20 7631 6900 E info@cipr.co.uk
W cipr.co.uk Twitter @cipr_uk

in decision making on service provision, service changes, service cuts and or new ways of delivering services.

In this instance, it is important for Standards to enable those working in UK Government Public Services and associated agencies and bodies to understand how this will affect the public.

As the private sector is likely to be involved in the provision and delivery of public services where data and artificial intelligence is incorporated, the CIPR AIinPR panel strongly recommends the UK Government's Artificial Intelligence and Public Standards applies to all private organisations/services and individuals providing data and/or AI consultancy, actual technologies, all or in part, and deployment.

We recommend the UK Government goes further and ask for a firm commitment that as part of all UK Government/public agency contracts awarded to provide data and artificial intelligence consultancy, technology solutions and deployment programmes - all private individuals, companies and groups must sign the Public Standards on Data and Artificial Intelligence as part of the conditions for legal contracts. This will provide the public with the assurances they rightly require on ethical use and decisions being made on their data and service provision.

This will also help support the upholding of trust by the public that UK Government Public Services are acting in the public interest and not causing any untoward harm to individuals, groups or society as a whole by not including the private sector in the UK Public Standards on Artificial Intelligence.

Transparency has to cross borders between public services and private services as well as between different organisations and agencies in the public sectors, particularly at a time when the drive is to provide more joined up services between appropriate agencies.

All parties providing services, support and consultancy, short-term and long-term, to the UK Government Public Sector must be expected to sign up to the Public Standards. Failing that, the CIPR AIinPR panel suggests this is grounds not to award contracts as this could become a business and trust risk. At a time when the government needs to be able to instil public trust in the new data-fuelled and AI-enabled services and decisions to the UK taxpayer, this is crucial.

The CIPR AIinPR Panel welcomes the devising and introduction of the UK Government Public Artificial Standards but calls for a clear commitment to the above points, as detailed, to be considered for incorporation.

85 Tottenham Court Road, London W1T 4TQ
T +44(0)20 7631 6900 E info@cipr.co.uk
W cipr.co.uk Twitter @cipr_uk
Company Incorporated by Royal Charter (England/Wales)
No. RC000801 VAT Registration Number 234 2948 59

SUBMISSION 17

Response of the Mission and Public Affairs Council of the Church of England (MPA) to The Committee on Standards in Public Life Artificial Intelligence and Public Standards Review

The Mission & Public Affairs Council of the Church of England (MPA) is the body responsible for overseeing research and comment on social and political issues on behalf of the Church. The Council comprises a representative group of bishops, clergy and lay people with interest and expertise in the relevant areas, and reports to the General Synod through the Archbishops' Council.

We welcome the opportunity to respond to The Committee on Standards in Public Life (CSPL) Artificial Intelligence (AI) and public standards review. Our response takes into account the discussion published in the CSPL: roundtable transcript 23 May 2019.

We offer this submission, based on MPA's work over the past three years to understand the challenges posed by AI at many levels so that the Church of England can respond openly and with an informed moral voice. We have been in dialogue with leading figures in AI in universities, with ethicists and with Parliamentarians, and have contributed through our staff team to conferences, round tables and colloquia in the UK and abroad, alongside leading figures in different relevant disciplines. MPA is a partner with the Universities of Durham and York in a major Templeton-funded programme, Equipping Religious Leaders in an Age of Science (ECLAS) which has made developments in AI a major theme of its work promoting greater confidence among church leaders in engaging with the realities, rather than the myths, surrounding science and religious belief. MPA is also a partner with the University of Bath (and some 30 industry partners) in setting up a Centre for Doctoral Training in Accountable, Responsible and Transparent Al (ART-AI) where our contribution will be specifically to explore the contribution to AI of ethics which go beyond utilitarian and consequentialist approaches. MPA is also working closely with the Bishop of Oxford in his extensive work on AI, including his membership of the APPG-AI and the Centre for Data, Ethics and Innovation (CDEI), and the Bishop's Parliamentary Researcher, working on these issues, is based in the MPA team.

Summary of Key Points

1. We endorse the Nolan Principles as providing a valuable framework for considering the role of AI in public life
2. We affirm the primacy of human agency as the moral focus of AI – Narrow AI involves delegated responsibility for tasks.
3. We note the dangers of "exceptionalism" in dealing with AI – the Principle of Parity requires equivalent outcomes online and offline.
4. Retention of public trust is a key concern in AI developments, as is the growth of Data (and other forms of) Inequality.

Suggestions and questions

1. AI is promoted both as a solution to the problem of boring repetitive work – and as the solution to work involving extreme complexity. There may be different kinds of moral implication in workplaces in the two cases. Do those differences affect the application of the Nolan Principles?
2. For the Principles to guide practice, it is important to consider questions of character and virtue among those seeking to live by the Principles. Professional bodies may have a role here.

The Seven Principles in a Common Good framework

As the CSPL review focusses primarily on the Seven Principles (the "Nolan Principles"[35]), this response begins by outlining the MPA's broader ethical perspective on AI in relation to those principles before going on to suggest some other, possibly fruitful, areas for attention. Whilst we recognise that the Seven Principles are designed to be broadly acceptable in a plural society, and thus are not expressed in specifically Christian or religious terms, they nonetheless cohere in many respects with the concept of the Common Good which is a central theme in Christian social thought. We would wish to evaluate the efficacy of the Seven Principles according to whether they support the following:

- the intrinsic value of human life (within the wider principle of the affirming of life)
- the protection of the vulnerable
- building a caring and cohesive society
- respect for the individual.

[35] https://www.gov.uk/government/publications/the-7-principles-of-public-life

The Church of England uses these four criteria in evaluating public policy decisions. Thus, within this framework, questions of efficiency and cost-effectiveness are subsidiary to questions of what types and uses of technology will promote best working environments and best practice in public bodies, with an ethical framework supportive of a healthy society. We endorse the arguments and use of language in this review which place those fundamental concerns at the centre of any regulatory decision-making processes – with or without AI 'in the loop'.

We note too that seeking to understand and anticipate the social impact of deploying AI technologies in the work of public regulatory bodies, context by context, task by task, is challenging. The way this task is approached from inception, through the consultation process, commissioning and design of AI technologies for specific contexts, and on to communication with the wider public, will have intended and unintended consequences in the complex organic reality that is human society. We recognise there is no perfect route but encourage extensive reflection and interdisciplinary consultation in order to anticipate at least some of the human costs as well as the benefits.

AI and inequality

Engagement with complex technologies can promote a sense of being disadvantaged both when being required to work with a new technology – that is, imposed from above – and when being the recipient of a formal decision, arrived at by technological means such as algorithms, which may appear to be above contradiction. We share the concern that AI technologies should be introduced in ways that counter potential inequalities both in knowledge and operational understanding and in the different levels of design power and accountability in an organization. This includes finding ways to avoid de-skilling and de-motivating the workforce. It also includes examining the way that formal decisions from public regulatory bodies are communicated to recipients of decisions. Ultimately, the very texture of public regulatory processes is being changed with a host of attendant challenges and questions to address. It would be helpful to know how the CSPL intends to address these questions, and with what bodies it proposes to share the burden of responding to the needs of those who are disadvantaged by changes in the workplace and the regulatory system.

Education and public communication are clearly essential and the way that the need for innovation – both in terms of technological change and responses to technological change – is both understood by senior decision makers and communicated to recipients of those decisions will be of critical importance.

The Mission and Public Affairs Council has a long history of challenging growing inequality in many areas of national life – including not only material inequality but inequalities of power and respect. The MPA staff team is about to begin work on the question of Digital Inequality, bringing together its work on the ethics of AI and other forms of inequality, and is likely, in due course, to seek the mind of the Church of England on ways to challenge the threat of Digital Inequality, through a debate at the General Synod, when its researches are more advanced.

Contextual sensitivity and awareness of impact on workforce

During the CSPL Roundtable discussion, we noted the repeated emphasis on the need to assess deployment of AI on a context by context basis. Some tasks can only be carried out by AI technologies, such is their scope and intricacy – national security related surveillance is one example. On the other hand, reflecting for instance both on the care of the workforce and on impact on productivity, it is important to ask what happens when routine tasks are performed by AI technologies and the worker is now presented only with 'difficult' tasks. One scenario assumes that a reduction of routine tasks will create more fulfilment for the worker in question and, therefore, more productivity. On the other hand, the absence of variation in levels of difficulty of task might, instead, be counter-productive, increasing stress and reducing effectiveness. Conversely, it is possible that people may be demotivated if too much of the difficult aspects of their roles are delegated to AI. It is worth asking whether there are areas of public life where the cost benefit ratio is not high enough to justify introduction of new technologies or only limited technological innovation. It will also be important to ask who will make that decision and on what basis.

AI and the Seven Principles: is re-evaluation of the Principles needed?

The review asks, "whether the use of AI requires a new understanding or formulation of the Seven Principles of Public Life?" This is not a question that can be answered with a simple Yes or No. On the one hand, we reject any appeal to legislative 'exceptionalism' that demands a distinction be drawn between ethical norms in the digital (online) and offline worlds. Such 'exceptionalism' would risk generating claims that these advanced new technologies require a special set of rules and that regulation should be minimised for the sake of ongoing innovation. 'Narrow AI' of the kind within the scope of this review is always an extension of human activity, not an external or ethically competing 'other'.[36] The 'advanced autonomy' the Committee's terms of reference describes is always a delegated autonomy. It is permanently framed by the human-directed goals of those deploying AI. The AI this Committee is considering should be treated, therefore, as a technological tool at the disposal of those working in the public sector instead of expanding the discussion to include potential 'general' AI and ethical demands that might reasonably conflict with those of an AGI's end user.

On the other hand, to add strength to this, the 'principle of parity' may be useful as articulated in the House of Lords Select Committee on Communications March 2019 report *Regulating in a Digital World*: 'We define the "principle of parity" to mean that regulation should seek to achieve equivalent outcomes online and offline'. Equally, consideration should be given to 'Ethical design', another principle emphasized in *Regulating in a Digital World*.[37] Placing uncritical trust in an "intelligent" technology – and this may include deference to machine decisions even when there is conflicting information – is a frequently reported phenomenon. Research is now taking place to establish the impact on human decision making of advances in the use of AI. In the context of this consultation, we argue that 'Ethical design' should include design elements that remind the user they are dealing with a machine with the capacity to get it wrong. Guidance about machine fallibility should also be widely available to the public – at a level that a young person can understand. Also protective against over-reliance on AI are clear systems of review and accountability within teams.

---

[36] Narrow' AI is commonly distinguished from Artificial General Intelligence (AGI). 'General' AI is considered to be equivalent to, or "better" than human intelligence, but remains a speculative technology at present.

[37] *Regulating in a Digital World.*

Technical questions : AI algorithms, reliability and ethical safety and accountability

We now move to the question of the technical reliability of AI-informed decision-making systems and accountability for these systems. Here, we continue to affirm the primary role of the human in decision making. Where AI algorithms are used, we would suggest that the Committee treat language referring to humans "in", "on", or "above" the AI loop as indicating discussion that is happening at a lower level than the whole system perspective. From this whole system perspective, the responsibility of human agents, and thus the Seven Principles, continue to apply. In part this is because the product of every AI algorithm is innately probabilistic:

> 'Most data sets in the internet age are collected under highly non-scientific conditions for someone else's purpose, and they are only incidentally useful for any other purpose. To get round that – or, just to know when you can't – you had better reckon honestly with your assumptions.'[38]

Even the most sophisticated and deeply multi-levelled recursive neural network (RNN) depends upon human direction for both its end goal and overarching methodology. Yet there are already stock examples from the medical profession showing how differences of initial assumption produce different conclusions even when evaluating the same question using exactly the same data set.[39] As the EU Data Protection Supervisor notes

> Algorithmic processes are 'only one among many possible ways of representing the world and its inhabitants and the digital transcription of behaviours and propensities is neither neutral nor exhaustive.' '[40]

The ethical challenge posed by so-called 'black box' algorithms should be evaluated in relation to both this lack of absolute objectivity at the algorithmic level and the need for a

---

[38] Nick Polson and James Scott, *AIQ: How Artificial Intelligence Works and how we can Harness its Power for a Better World, Penguin,* 2018, p.228.
[39] C.R. Cardwell et al., "Exposure to Oral Biophosphonates and Risk of Esophageal Cancer'" *JAMA 304,* no. 6 (August 11, 2010): 657-63; J. Green et al., "Oral Biosphosphonates and Risk of Cancer of Oesophagus, Stomach, and Colorectum: Case-Control Analysis Within a UK Primary Care Cohort," *BMJ* 2010;341:c4444.
[40] European Data Protection Supervisor (EDPS) Report, 2018, pp11-12. https://edps.europa.eu/sites/edp/files/publication/18-01-25_eag_report_en.pdf, accessed July 3rd, 2019.

whole-system perspective. The Committee has already received testimony about the potential challenges AI presents for explainability at the algorithmic level.[41] Yet the All Party Parliamentary Group for Artificial Intelligence (APPG AI) has also received testimony that the claims around opacity of 'black box' outputs may reflect issues of process governance (explanatory competence) rather than inherent explainability and that black box accounts of decision making threaten public trust.[42] Safety and ethical legitimacy by design depends upon continuous and connected human oversight at the whole system level from procurement to deployment. It is this holistic Public Sector process that should comply with the Seven Principles.[43]

What happens when these Principles fail?

Trust depends not only on promoting high standards of personal conduct, but also on repairing lapses when they happen. Consequently, we urge the committee to structure oversight of the Seven Principles in a way that promotes public trust in the whole system.

As Luciano Floridi argues…

> Public acceptance and adoption of digital technologies, including artificial intelligence, will occur only if the benefits are seen as meaningful and risks as potential, yet preventable, or minimizable, or at least something against which one can be protected. These attitudes will depend in turn on public engagement with the development of technologies, openness about how they operate, and understandable, widely accessible mechanisms of regulation and redress.[44]

---

[41] https://www.gov.uk/government/publications/artificial-intelligence-and-public-standards-roundtable-transcripts, accessed July 4th 2019.
[42] AI All Party Parliamentary Group, Evidence Meeting 3, May 2019. Testimony by Dr. Matthew Howard, Director of Artificial Intelligence and Cognitive Analytics, Deloitte (Cf., 1:08:30ff) https://www.appg-ai.org/evidence/2019-evidence-meeting-3-implementation-full-video/
[43] For a good practical example, see the National Endowment for Science, Technology and the Arts (NESTA) template: 20 questions for evaluating the use of AI in the Public Sector. https://www.nesta.org.uk/blog/20-questions-public-sector-orgs-algorithms/, accessed July 8th, 2019. Government of Canada, Algorithmic Impact Assessment https://www.canada.ca/en/government/system/digital-government/modern-emerging-technologies/responsible-use-ai/algorithmic-impact-assessment.html, accessed July 3rd, 2019.
[44] Floridi L. 2018 Soft Ethics, the Governance of the digital and General Data Protection Regulation. *Phil Trans. R. Soc. A* ***376****:* 2018008. http://dx.doi.org/10.1098/rsta.2018.0081, accessed 8th July 2019.

The contribution of character – Human virtue and standards in AI.

The very fact that questions around AI are of interest to a body considering standards in public life helps emphasise the point we have made already that the human agency behind the algorithm – however diffuse and remote – must always be kept in focus. Thus, whilst it is inevitable that the primary focus on the public good will be on regulation or the establishment of fundamental principles of conduct, we would like to introduce into the discussion the centrality of character and the formation of virtuous instincts in those who bear great moral responsibilities (whether they are aware of the magnitude of their responsibilities or not). In other contexts, professional self-regulation remains an effective mechanism for enforcing and policing a robust professional ethic. In medicine, the law and some other fields, violation of a professional code of practice, which may be set out not only in terms of rules but in terms of behavioural expectations, leads to discipline by one's peers, and ultimately expulsion from the profession and inability to continue to practice. As we argued some years ago in our submission to the Commission on Banking Standards, it was partly the loss of such a body of professional wisdom and the structures that went with it which contributed to the breakdown of trustworthy behaviour in the banking industry. We recommended then that the restoration of a professional body and a professional ethic would enable good practice to be policed, not just by regulation but by shame which, we believe, could be the basis of a very powerful and internalised sense of when ethical boundaries are at risk of being violated.

The world of AI is not strictly comparable to banking or other professions, not least in the much more remote relationships between those who bear responsibility for the design and end use of algorithms, those who commission that design and determine the context in which it will be applied, and those most affected by the deployment of algorithms of which they may never be aware. Nevertheless, we suggest that, at different stages in that chain, the task of identifying where responsibilities lie could be accompanied by consideration of how professionalism and professional ethics can be informed by culture as well as by regulation. The Seven Principles provide a useful starting point here. But making them "stick" relies not only on a "tick box" approach to their application but also on their centrality in the professional cultures around AI design, application and deployment. The central moral question might be posed as "what does it mean to be a good AI technician or user of AI?"

where the word “good” connotes a set of practices and cultures and not just externalities such as profit creation or awe-inspiring design factors (although these are also part of the picture).

We are aware that much more work would need to be done to make this suggestion a practical and effective concrete provision, but neglect of the question of character in the context of those with responsibility for AI developments risks colluding with the impression that AI is an autonomous other to which questions of human agency and responsibility simply do not apply. As we have suggested above, the tendency already for people to trust AI-generated conclusions above their own reasoning or assessment of evidence is starting to place AI developments beyond the reach of moral enquiry and thus, of public standards. A focus on the professional ethos of AI-related professions would help counter this by placing personal responsibility at the centre of the debate. Moreover, the focus on character and virtue in professional standards would cohere with important trends in moral philosophy (including but not confined to Christian ethics) over the last 40 or so years.

Mr Mark Sheard
Chair, Mission and Public Affairs Council
of the Archbishops’ Council of the Church of England.

July 2019

SUBMISSION 18

Centre for Data Ethics and Innovation

Response to the Committee on Standards in Public Life's AI & Public Standards Review

31 July 2019

1. The CDEI's developing approach to ethical and governance principles for data-driven technology

In recent years a plethora of governance and ethical frameworks have been developed to help guide ethical development, deployment and governance of new technologies. We welcome the recent commitment by 42 countries, including the UK, to adopt the OECD human-centred Principles on Artificial Intelligence[1]. Some of the other frameworks we have drawn on to inform our work include: the European Commission High Level Working Group on AI[2], the UK Parliament Select Committee on Artificial Intelligence's Report, AI in the UK[3], and the UK Government's Guide to using Artificial Intelligence in the Public Sector[4].

We have identified - among these sets of principles - three areas to frame our thinking and to underpin our work:

1. ***Societal values:*** These values set out how we want the world to be and include notions such as fairness, autonomy and social cohesion. A vision of what constitutes a healthy society in the era of data-driven technology.
2. ***Principles for good governance:*** These principles set out how we seek to arrive at outcomes that are consistent with our societal values. This includes principles such as proportionality, accountability, and transparency.
3. ***Levers for achieving good governance:*** Practical frameworks and mechanisms for how to achieve good governance. They include monitoring, enforcement and compliance. They may be a combination of (1) legislative and regulatory measures, (2) soft governance such as standards, codes of practice and corporate governance, and (3) technical approaches such as privacy enhancing technologies or design approaches for explainability.

These are not unique to AI; they are principles and levers of general application. There are different ways of grouping and describing them, but there is a high degree of convergence about the principles that will help us govern AI[5] and the are the same principles that have served us well historically.

[1] https://www.oecd.org/science/forty-two-countries-adopt-new-oecd-principles-on-artificial-intelligence.htm
[2] https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai
[3] https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf
[4] https://www.gov.uk/government/collections/a-guide-to-using-artificial-intelligence-in-the-public-sector
[5] AI4People—An ethical framework for a good AI society : Opportunities, risks, principles, and recommendations . Floridi, Luciano et. al in Minds and machines, Vol. 28, No. 4, 01.12.2018, p. 689-707

Using this terminology, the seven Nolan Principles are principles of good governance and are closely related to the various principles that inform our work. AI and algorithmic decision-making tools (ADMTs) create challenges at all three levels of governance.

1. **Societal values:** Societal values are usually expressed in contestable terms such as 'fairness' and 'discrimination' that are then further defined in statute, case law and regulation. Big data creates an unprecedented ability to measure issues such as variation in decision making or discrimination between different groups. This creates pressure to more explicitly define trade-offs between competing and (sometimes incompatible) definitions of concepts such as fairness.

2. **Principles for good governance:** Principles of good governance face new questions in a data driven world. For example, the principle that decisions will be objective and based on evidence implies an understanding of what constitutes an appropriate level of evidence. In a world of rich and deep data, evidence is much more contestable. Organisations need to balance an expectation that evidence is contestable with the need to reach a decision in a reasonable time frame and with constrained resources. Similar issues apply to principles such as accountability and openness.

3. **Levers for achieving good governance:** This is perhaps the area where there is the greatest need for the development of guidance appropriate to a data driven world. Many of our traditional mechanisms of governance, oversight and regulation have been built in an analogue world. The techniques and capabilities required to understand and govern the operation of automated systems differ from those used to govern human systems and are very context specific.

We are mindful of the significant benefits that AI can bring to the UK - including through its adoption by the public sector. Guidance and requirements for public servants should be proportional to the stakes involved and should facilitate the adoption of trustworthy AI that can make our public sector more effective and efficient.

2. Ensuring the implementation of standards

Resources for public servants already exist, including the Data Ethics Framework and the Office for AI's Guide to Using AI in the Public Sector. But further work is needed to expand on this and promote the adoption and implementation of measures that will facilitate the continued application of the Nolan Principles in a public sector characterised by the increasing use of AI.

An assessment of the level and spread of uptake of existing guidance across the public sector may be useful in indicating whether additional obligations or incentives are required to maximise the impact of work to articulate how principles apply to the use of AI.

3. CSPL's Challenge 1 - ***"the balance of responsibility between humans and AI for decisions and processes in the public sector must be clearly defined."***

Humans must be ultimately responsible for decisions made by any system. In this sense, there is no 'balance' to be struck between humans and AI. However, there will be variation in the degree of responsibility *between* those humans that are accountable for or interact with the system.

Where an ADMT has had a role in a public sector decision, there will be a number of humans involved. Putting the humans involved into broad categories can help us identify the responsibility they have and the standard of behaviour we can expect from them. For example:

Human	***User***	***Commissioner***	***Reviewer or Ultimate Authority***
Role and responsibilities	● Enter data ● Review recommendation and act on it ● Communicate final decision	● Decide to introduce ADMT ● Provide training data for development ● Develop or procure the system ● Integrate it into decision-making processes	*Specific to sector or public body, e.g. Minister or CEO*
Expected standard	● Follow user guidelines ● Act appropriately and use own judgment (within realistic expectations)	● Specify appropriate goals for the ADMT ● Collect and provide appropriate training data ● Follow good procurement practice ● Develop appropriate training for users ● Assess the expected and actual impact of the introduction of the ADMT	Set organisational culture in line with - and monitor implementation of - agreed principles

Good governance will require for each use case, a specific understanding of the appropriate division of responsibilities. For example, consider the use of facial recognition technology and automatic passport screening to determine whether or not someone can enter the country at an airport. In this scenario, there may be no human oversight of individual decisions, which increases the responsibility on the commissioners and reviewers to ensure error rates - both false positive and false negative - are within acceptable bounds. In contrast, if a doctor uses an algorithm to inform a diagnosis, their professional responsibility for making an accurate diagnosis is not altered. But in some cases this responsibility imposes an obligation to use an algorithm, or even a presumption to follow its advice; while in others, an algorithm might be something that the doctor uses at their own risk.

As the table indicates, we may be most concerned with the accountability and conduct of the public servants commissioning, designing and introducing ADMTs into decision-making processes. In many if not most cases the lines of accountability will already be clear, with specific individuals having overall responsibility for the delivery of policies and services. We should expect high standards of conduct from senior managers with overall responsibility for decision systems - and therefore for the role of ADMTs within them. We should also work to ensure they have the guidance and tools they need to maintain those standards.

4. CSPL's Challenge 2 - *"AI must not create, amplify or reflect real world demographic bias."*

Defining bias in the context of decision-making is challenging. In general usage, when we describe a decision as biased, what we mean is that it is not only skewed, but skewed in a way which is unfair. As the volume and variety of data used to inform decisions increases, and the algorithms used to interpret the data become more complex, concerns are growing that without proper oversight, algorithms risk entrenching and potentially worsening bias. Our review into bias in algorithmic decision-making is exploring bias in two public service sectors: policing and local government (as well as in the private sector in recruitment and finance).

The use of ADMTs also has the potential to improve the quality of decision-making by increasing the speed and accuracy with which decisions are made. If designed well, they can reduce human bias in decision-making processes. However, this requires us to make choices in an accountable and transparent way about the design of algorithms and the degree to which the factors that inform algorithmic outputs are legitimate and proportionate.

Given the challenges of determining what is 'fair', it is important to understand the status quo prior to the introduction of data-driven technology in a public sector. It is unlikely that any decision-making system will be completely free of potential bias. This understanding of historical context is relevant to assess the level of potential harm or benefit to be derived from the new approach. Whilst bias itself may not be unacceptable in principle, what is

likely unacceptable is *not to know* how a decision system is biased, and not to have openly justified that bias by reference to principles of conduct as a whole.

Our review seeks to answer three questions, which may be informative in considering how to support public servants to uphold the Nolan Principles:

1. **Data: Do public servants have access to the data they require to adequately identify and mitigate bias?**

Data itself is often the source of bias but, at the same time, it is a core element of tackling the issue. Our early research on the use of predictive analytical tools in policing suggests that one of the key issues with regards to the use of this technology is potential bias embedded in historic datasets.

It is common practice to avoid using data on protected characteristics (or proxies for those characteristics) as inputs into decision-making algorithms. However, understanding the distribution of protected characteristics among the individuals affected by a decision is necessary to identify biased impact. This tension between the need to create algorithms which are blind to protected characteristics while also checking for bias against those same characteristics creates a challenge for organisations seeking to use data responsibly.

2. **Tools and techniques: What statistical and technical solutions are available now or will be required in future to identify and mitigate bias and which represent best practice?**

As the systems which inform decision-making become increasingly complex and data intensive, it can be difficult to establish where bias has originated. Organisations using decision-making algorithms have an interest in evaluating potential unintended biases emerging from these systems, creating a need for bias identification tools and techniques.

Approaches to evaluating decision-making algorithms for bias are beginning to be proposed, either by academics in literature or by interested groups or companies as products or services. There is limited understanding of the full range of these approaches. Across the board there appears to be a lack of clarity over the relative strengths and weaknesses of these tools.

3. **Governance: Who should be responsible for governing, auditing and assuring algorithmic decision-making systems?**

Data gathering and analytical tools can help to understand the presence of bias in decision-making, but this is only a first step. We must subsequently decide how far to mitigate bias and how we should govern our approach to doing so. Humans are often trusted to make these trade-offs without having to explicitly state how much weight they have put on different considerations. Algorithms are different. They must be programmed

to make trade-offs according to explicit rules. This requires a different approach to accountability.

Public servants are likely to face significant trade-offs between different kinds of fairness, and between fairness and accuracy. There is currently limited guidance and a lack of consensus about how to make these choices or even how to have constructive and open conversations about them. These choices are likely to be highly context specific.

New functions and actors, such as independent bodies with a specific mandate, may also be required to independently verify claims made by public bodies about how their algorithms operate.

Our approach in the context of policing focuses in particular on the proper conduct of trials of AI for crime prevention and detection. Well-planned trials, with results subject to the appropriate scrutiny, should help identify whether the ADMT(s) being introduced are within our tolerance for bias in public sector decision-making. We are developing a code of practice for predictive policing trials, which will be published in draft in autumn 2019.

5. CSPL's Challenge 3 - ***"AI must be used in the public interest. It must be used to serve a clear public need and not to grant unnecessary or disproportionate powers to the state and public bodies."***

Impact assessments utility in balancing public interests

There will most often be at least an arguable public interest in deploying an ADMT. In many cases this may be as simple and compelling as making financial efficiency-savings. The greater challenge is deciding how to balance *competing* public interests, or the public interest against individual disbenefits.

It will not always be consistent with the Nolan Principles for the public body introducing an ADMT to be the sole arbiter of how to balance those competing interests. The principles of Objectivity, Accountability, Honest and Openness will require that these trade-offs be supported by:

- Tools and frameworks to systematically assess the likely and demonstrated impact of the introduction of automated decisions;
- Transparency for those impact assessments and accountability in decision-making; and
- Open and informed public debate.

Impact assessment of ADMTs in the public sector may range from informal, voluntary self-assessment through to legally-mandated third party assessment by a nominated body.

The Canadian model is at the more formal end of the spectrum, but remains relatively untested as yet. Whatever type of system is introduced, public servants must be incentivised in some way to carry out impact assessments and act upon their results, without being constrained from adopting beneficial innovation.

Facial recognition technology

The use of live facial recognition technology by the police is one area in which particular concerns have been raised about disproportionate state power over individuals. We are investigating this issue and would be pleased to share our initial findings with the Committee in September 2019.

6. CSPL's Challenge 4 - *"AI decisions and processes must be transparent."*

Transparency about the use of AI in the public sector

A first order problem faced in putting the Nolan Principles around transparency into practice for AI is that we do not have systematic knowledge of where ADMTs are being used in the public sector. Indeed we lack a good working definition of AI for the public sector and there is no authority mandated to monitor the extent to which AI is being introduced into public life. This is and will continue to be a barrier to full transparency and accountability for public sector AI decisions[6].

We note the recommendation by the Law Society that a national register of ADMTs in use in criminal justice be established. Subject to appropriate exceptions, thresholds and safeguards, this would appear to support the Nolan Principles and would facilitate impact assessment of public sector ADMTs. Such a register may be appropriate in other parts of the public sector.

Transparency for individual decisions involving ADMTs

Claims about what is technically (im)possible should be treated with caution. Our engagement with industry to date suggests that, if a degree of explainability is made a priority from the outset by its commissioner, it can be built in.

Proportionality is an important principle here. There is understandable concern in some parts of the AI sector that requirements to explain comprehensively how every data point in a model influenced a specific outcome will limit the accuracy and efficiency of ADMTs. However, a proportionate requirement to provide an adequate explanation, in a manner that facilitates follow-up, may not conflict with innovation, efficiency or accuracy.

[6] It is notable that lack of transparency about where ADMTs are in use has been a major barrier to algorithmic impact assessments in New York State.

Transparency around individual decisions in the public sector needs to be linked to clear procedures for challenge, appeal or redress where standards have not been upheld. Care should be taken to tailor the degree of transparency to be useful to the individual or entity impacted by a decision - raw source code is unlikely to be useful to an ordinary citizen affected by a public decision.

7. The CDEI's work and ongoing role

Our work on algorithmic bias, facial recognition, and other issues of potential relevance to the Committee is ongoing, and we would be pleased to discuss the results of that work with the Committee as they become available.

We believe CDEI has a role to play in anticipating challenges posed by AI to the Nolan Principles, and in developing guidance and tools to help public servants put the principles into practice in an AI context. In particular, CDEI has a mandate to provide expert advice to regulators and to help coordinate the regulatory landscape. CDEI will be most effective in maintaining public standards in relation to AI as a statutory body, independent of central government policy-making, and with stable and reliable funding. We are considering what specific statutory form, functions and powersCDEI needs in the long-term to be most effective, and would welcome further discussion of this with the Committee.

[redacted]

[redacted]

[redacted]

SUBMISSION 19

[redacted]

In addition to Simon's oral contribution we wanted to send you some written material which expands on what we are doing in this area.

Firstly, you've probably seen our tech and innovation page which has general background to AI and data protection.

https://ico.org.uk/about-the-ico/what-we-do/tech-and-innovation/

On our ExplAIn project, the latest material is the Interim Report: https://ico.org.uk/media/about-the-ico/documents/2615039/project-explain-20190603.pdf

In addition, this page https://ico.org.uk/about-the-ico/research-and-reports/project-explain-interim-report/ links to our blog post about the report and to materials about the citizens juries part of the research from the University of Manchester.

Material on our initial views from the AI Audit Framework and our plans for this work –

https://ai-auditingframework.blogspot.com/2019/03/simon-mcdougall-director-for-technology.html

https://ai-auditingframework.blogspot.com/2019/03/an-overview-of-auditing-framework-for_26.html

https://ai-auditingframework.blogspot.com/2019/07/developing-ico-ai-auditing-framework.html

[redacted]

[redacted]

I hope this useful for the Committee. Please feel free to ask any questions, or request clarification on anything.

[redacted]

[redacted]

[redacted]

[redacted]

SUBMISSION 20

Submission to Committee for Standards in Public Life
Review into Artificial Intelligence and Public Standards

From Professor Karen Yeung
Interdisciplinary Professorial Fellow in Law, Ethics & Informatics
Birmingham Law School & School of Computer Science

Executive Summary

The 'Public Interest' requires that government use of AI-driven systems must respect foundational principles upon which constitutional democratic political orders rest:

1. The rule of law;
2. Respect for human rights; and
3. The preservation and maintenance of democracy.

In addition, principles of good public administration require that the use of these systems should be

4. Effective in promoting the achievement of legitimate public interest objectives

Detailed analysis

Each of these four principles is elaborated upon below, and illustrated primarily with respect to the increasing use of facial recognition technologies by the police.

1. The Rule of Law

Core principle: Public sector use of algorithmic systems based rest on a clear, transparent, explicit lawful basis for its use. Within the British legal framework public office-holders have no power other than that lawfully conferred upon them. They have a fundamental duty to observe the rule of law – this is inherent in obligations of accountability (although not explicitly stated)

- The need for explicit lawful basis is of paramount importance given that these technologes interfere with individual rights and freedoms, and do so (1) at scale (2) in real time (3) operating in highly opaque ways which are extremely difficult (if not practically impossible) for affected individuals to understand, let alone (4) challenge and contest.

- Not adequate to employ technical legal arguments to 'cobble together' an 'implicit' lawful basis, given that power, scale and intrusiveness of these technologies create serious threats to the rights and freedoms of individuals, and to the collective foundations or our democratic freedoms. Yet in current litigation challenging use of facial recogition tech by UK police (Liberty, Big Brother Watch) the Police claim that there IS a lawful basis, by seeking to cobble together a patchwork of legal sources which they claim provides implicit legal authority.

- The way in which facial recognition technology ('FRT') has been tested and used in live e environments is likely to be in breach of data protection law:

  Because FRT and other biometric technologies entail 'sensitive processing' under the EU Law Enforcement Directive (and incorproated into UK law under the Data Protection Act 2018 Part III) it is subject to very stringent conditions on use. Current use of FRT is very unlikely to meet these conditions, and hence likely to be unlawful under the Data Protection Act 2018, Part III (the domestic implementation of the EU Law Enforcement Directive)

  - Facial recognition for law enforcement purposes constitutes "sensitive processing" under the DPA 2018, as it is biometric data processed for the purpose of uniquely identifying a natural person.

  - When not relying on consent of the individual, processing this data must be

    (1) strictly necessary for a law enforcement purpose
    (2) meet a condition in Schedule 8 of the DPA 2018, and
    (3) be accompanied by an appropriate policy document.

  - It is highly unlikely that facial recognition at scale can meet a test of strict necessity, particularly given its highly unproven nature (See P Fussey & D Murray (2019) *Independent Report on the London Metropolitan Police Service's Trial of Live Facial Recognition Technology*, University of Essex. Hereafter 'Independent FRT Review') and observes that there is no evidence of effective promotion of law enforcement goals).

2. Respect for human rights

Core principle: Within constitutional democratic states, actions by public authorities must not violate human rights.

2.1 Facial recognition technologies are especially troubling and dangerous

These technologies pose serious threats to both human rights and democracy, fundamentally reversing the presumption of individual liberty that has, for centuries, been situated at the heart of British constitutional and political culture.

- As Liberty (2019) '*Policing By Machine*' report states:

  > "a dangerous emerging narrative requires us to justify our desire for privacy, rather than requiring the state – including the police – to provide a sound legal basis for interference." (p 26)

- Facial recognition technologies generate serious dangers of disproportionate use (eg using live facial recogntion to catch individuals suspected of petty and minor offences), clearly violating human rights law principles while seriously threatens our culture of presumptive individual liberty and freedom:

- Eg Independent FRT Review (2019) by Essex University scholars of live facial recognition technology ('FRT') was trialled in operational settings across the UK in which FRT were applied to

detect individuals on a 'Watchlist' used for London MPS live facial recognition trial, which independent review concluded was unlawful in violation of human rights standards.
"Issues to do with the accuracy of the watchlist played out when individuals were stopped on the basis of outdated information. On occasion, individuals were flagged by the LFR technology in relation to a serious offence, but this had already been dealt with by the criminal justice system. However, they were wanted in relation to more minor offences and were arrested accordingly. It is unlikely this lesser offence would have been sufficiently serious to be included in the initial watchlist. This raises additional concerns when LFR is deployed on a necessity calculation intended to address serious crime but is then also used for more minor offences." (p 11)

"implicit legal authorisation claimed by the MPS for the use of LFR - coupled with the absence of publicly available, clear, online guidance – is likely inadequate when compared with the 'in accordance with the law' requirement established under human rights law."

2.2 Police databases contain unlawfully retained data

- FRT seeks out matches against images stored in police databases.

- Yet existing police data bases are likely to contain unlawfully retained images (including images taken of individuals brought into custody but never charged or convicted of an offence), and the way in which this data is stored and processed appears to be in breach of legal reqirements under the Data Protection Act 2018

- Eg. Custody photographs – retention and inclusion in Police National Database despite clear breach of PACE (per High Court ruling in 2012). See Law Society (2019) *Algorithms in the Criminal Justice System* which noted that in 2012, the High Court ruled in favour of two individuals who challenged the Metropolitan Police Service for retaining custody photographs taken under PACE (169). This ruling stated that the "existing policy concerning the retention of custody photographs… is unlawful", and the police were given a "reasonable further period" for revising this policy – a period which should "be measured in months, not years"(170). The report also observed that:

"Significant contention has surrounded the seeming non-implementation of the ruling in this act.

- The Biometrics Commissioner was established in 2012, but the mandate of the role did not contain photographs.(171) Nevertheless, the annual reports from the Office of the Biometrics Commissioner consistently challenge photographs and the regulatory inaction and governance vacuums that are perceived to surround their use. (172)

- The House of Commons Science and Technology Committee has raised also this issue in their report on biometric data and technologies, noting that they are: "…particularly concerned to hear that the police are uploading photographs taken in custody, including images of people not subsequently charged with, or convicted of, a crime, to the Police National Database and applying facial recognition software. Although the High Court ruled in 2012 that existing policy concerning the retention of custody photograph by the police was "unlawful", this gap in the legislation has persisted."(173)

- In February 2017, the government gave non-convicted individuals the right to ask police forces to delete their images from custody image database. A year later, 67 applications for deletion had been made, with only 34 successful.(174) This suggests that the current method for storing and deleting custody images is ineffective, and the approach stands in contrast to the millions of photographs stored in the Police National Database.

- Furthermore, the fact that photographs are not labelled with the status of the individuals within it.

This is unlawful under the Data Protection Act 2018 which requires in section 38(3) that:

> "In processing personal data for any of the law enforcement purposes, a clear distinction must, where relevant and as far as possible, be made between personal data relating to different categories of data subject, such as–
>
> (a) persons suspected of having committed or being about to commit a criminal offence;
> (b) persons convicted of a criminal offence;
> (c) persons who are or may be victims of a criminal offence;
> (d) witnesses or other persons with information about offences."

It appears the photographs used in the Police databases (and hence in the MPS Watchlist for the FCT trials) were not been distinguished in this way, which is likely to constitute an infringement of this legal requirement.

3. Democratic governance (transparency, accountability and public consultation)

3.1 Transparency and accountability

- There is a serious lack of transparency and concomitant lack of accountability about how the police and other law enforcement agencies are *already* using these technologies.

- Very little attempt to inform the public of how these technologies are in use, or the contemplated future uses. Accordingly, there has been very little public debate and discussion about whether these technologies should be used at all, let alone scrutiny of their policies and operations. Instead, there appears to be a general culture of secrecy within the public sector concerning the current and proposed use of these controversial and highly intrusive technologies

<u>Evidence and examples:</u>

Eg MPS live facial recognition trials started in 2016, but no information about these trials was publicly available until mid 2018

> "The first MPS documents obtained by the authors which provides information on the use of LFR technology are dated 23 and 25 July 2018.(213) From information available online, the MPS' LFR trial website – which contains information relating to the trial process, the legal mandate, and the data protection impact assessment –appears to have been created on 15 July 2018. (214) This indicates that no detailed information was available to the public prior to 15 July 2018 at the earliest. As the first trial was conducted in August 2016, and a total of five trials were conducted prior to 15 July 2018, this is clearly of concern with respect to the public availability of information, as required by the Surveillance Camera Commissioner's Code of Practice." (Independent FRT Evaluation (2019) at p 63)

Although authors of report state that the MPS in good faith published the info on the website and not by a desire to hide what they were doing, or keep disclosures to the regulatory minimum (p 63)

Law Society (2019) *Algorithms in the Criminal Justice System* observes

- "When decision systems are introduced into public contexts such as criminal justice, it is important they are subject to the scrutiny expected in a democratic society. Algorithmic systems have been

criticised on this front, as when developed in secretive circumstances or outsourced to private entities, they can be construed as rule making not subject to appropriate procedural safeguards or societal oversight.(71) Where algorithms are deployed by private sector organisations directly, freedom of information law has limited current applicability.(73))

- "It is notable that the Home Office does not routinely make information about facial recognition using the Police National Database public, with information only appearing on request in relation to parliamentary questions or in an *ad hoc* manner to offices such as that of the Biometrics Commissioner.(191) "

3.2 Serious gaps in current governance regime

- Law Society (2019) emphasises the need for oversight of algorithms in criminal justice to ensure accountability, particularly for FRT

"Few provisions currently support civil society organisations or forms of collective oversight of algorithmic systems directly, leaving a significant accountability gap in need of remedy (72). Furthermore, it is unclear whether civil society organisations have the capacity to engage in meaningful oversight, particularly given the rapidity with which different systems are being deployed across the sector and across the world.(74)"

…Many public bodies have reported on the lack of governance of police facial recognition.

- The Biometrics and Forensics Ethics Group, an advisory non-departmental public body sponsored by the Home Office, reported on police use of facial recognition systems and potential ethical frameworks, noting in particular the lack of independent oversight and governance.(186)

- The House of Commons Science and Technology Select Committee raised similar concerns.(187)

- The Information Commissioner launched an inquiry in December 2018 into police use of facial recognition technology,(188) having previously written a public-facing blog noting "how facial recognition technology is used in public spaces can be particularly intrusive" and that she was "deeply concerned about the absence of national level coordination in assessing the privacy risks and a comprehensive governance framework".(189)

- The Biometrics Commissioner has been consistently critical of the state of facial recognition, despite it falling outside of the role's statutory remit (190).

- The Law Society (2019) recommends a strengthened Biometrics Commissioner, both in terms of statutory responsibilities and resourcing.

Sub-Recommendation 5.3 Biometrics Commissioner – The scrutiny powers, resources, and consultation role of the Biometrics Commissioner should be strengthened, and the scope of the Commissioner broadened and regularly reviewed.

4. Are algorithmic decision-making systems *effective* in promoting legitimate public policy goals? (required by principles of good public administration)

Core principle: Algorithmic predictions must be approached with *great* care, even when they are used to promote legitimate public policy objectives motivated by a desire to promote the public interest.

- The turn to AI-driven decision-making in the public sector rests on an unexamined belief that the algorithmic feedback systems that underpin the success of Google, Amazon, Facebook and the like in generating accurate predictions of individual behaviour can be readily replicated in the public sector.

- Because the use of predictive algorithmic systems to inform decision-making rely upon past data as the basis for 'ground truth' – they are designed to, and inevitably replicate, past behaviours and structures. So there is an inherent propensity to ossification of existing social practices and structures, and thus to perpetuate historic injustices and discrimination, including replicating the disadvantageous treatment of various groups (women, ethnic minorities, etc). Eg 2015 study of the use of Google's AdFisher, women were shown high paying job ads compared to men on a 1:6 ratio. But they had no way of knowing this was occurring.

- But their use in the criminal justice system is <u>particularly dangerous</u> due to a widely held but often mistaken belief that of AI driven techniques offer accurate predictions about future behaviour.

- Why mistaken? Because the criminal justice context is NOT analogous to commercial advertising and other on-line content recommendation systems in which the user provides instant feedback via her click-through behaviour (does she click on the ad served up or not?) and these feedback loops operate in a direct and continuous cycle.

- Compare criminal justice algorithms to predict recidivism risk, such as the HART algorithm being used by Durham police force. In the criminal justice context, we do not have any reliable and accurate data set that can serve as 'ground truth.'

- HART algorithm being used by Durham polic force is based on historic arrest data ('custody events') which is treated as 'ground truth'. But arrest data does NOT provide an accurate prediction of the crime taking place in the relevant area. It merely provides data on those arrested by police for suspected criminal activity. It is NOT an indicator of

    - whether those individuals actually committed a crime (ie not the same as conviction data)

    - it does not contain data on crime that is unreported, nor

    - nor does it include record data of crimes committed for which no arrests have been made

- Accordingly, any algorithm based on historic arrest data will merely reflect social biases and prejudices associated with the making of arrests.

- Because a preponderance of low income, poorly educated individuals arrested for street crime are much more likely to be arrested than high income, white collar criminals who commit crimes of dishonesty which may generate orders vastly greater harm (at least in financial terms) compared with minor petty theft – yet these white collar crimes are not represented in the data set, and hence they are excluded from the algorithmic calcuations. Hence the algorithmic system will not only generates inaccurate predictions about the likelihood, distribution and character of 'future crime, but the predictions it generates will be seriously biased and thus generate discriminatory effects, because they bear no proven relationship with *actual crime patterns.*

Liberty sums up these effect in the following terms

- Effect of reliance on PP programs is beyond increased profiling and inefficient resource use – but to deepen divisions between the police and local communities, and these divisions can increase crime

- Portrayed as discovering new and useful trends, yet many will simply reinstate existing patterns of exclusion and inequality. They are NOT offering new insight into crime with greater utility for public safety: instead they lead to a cycle of self-fulfilling prophecies that unfairly target over-policed communities and undermine community relations with the police.

- Decisions of this kind are too important to hand over to machines – the risks to our civil liberties are too great. Any policing approach which relies on a combination of predictions, profiling and the idea of 'pre-criminality' puts our rights at risk. In a democracy that should value policing by consent, it's time to draw some red lines about how we want our communities to be policed (p 9)

- Council for Data Ethics and Innovation (CDEI) is undertaking work on algorithmic 'bias'. This is fine, as far as it goes, but 'bias' has been rather narrowly construed in the academic and public debates, couched in terms of its discriminatory effects on groups, particularly vulnerable and other marginalised groups. This IS important, but the problem of 'discrimination ' is only one kind of pernicious threats arising from the use of these data-driven approaches to prediction in the criminal justice system.

Recommendations

(1) We need **national leadership** (principle of LEADERSHIP)

- It is completely inadequate to leave it to local police forces and individual public sector agencies to 'make it up as they go along': not fair on them, and not fair on individuals and communities.

- Problematic that Council for Data Ethics and Innovation (CDEI) is located within DCMS and yet it is now formulating guidelines for criminal justice system.

- Leadership needs to be located at Cabinet Office level.

- For the criminal justice system and other law enforcement agencies, we need (1) national set of binding standards and guidelines and (2) National Public Register of algorithms in public sector use, per Law Society (2019) recommendations

    - Sub-recommendation 3.3: Code of Practice for Algorithmic Systems in Criminal Justice – The Government should request and resource the Information Commissioner to create a code of practice for algorithmic systems in criminal justice under the Data Protection Act 2018 s128(1).

    - Sub-Recommendation 1.7 National Register of Algorithmic Systems A register of algorithmic systems in criminal justice should be created, including those not using personal data, alongside standardised metadata concerning both their characteristics, such as transparency and discrimination audits and relevant standard operating procedures, and the datasets used to train and test them. Leadership of this could be taken by the Centre for Data Ethics and Innovation, as the Centre matures, in an open consultation procedure considering the criteria and thresholds for systems included in this register.

(2) The turn to AI-driven decision making in the public sector must be used with *great* care and appropriate scepticism, particularly when deployed within the criminal justice system (principle of OBJECTIVITY)

(3) An explicit lawful basis should be required for ALL algorithmic systems in use in the public sector (principle of ACCOUNTABILITY and INTEGRITY)

- It is vital that we demand that *all* algorithmic systems in use in the public sector are grounded on a publicly stated and explicit lawful basis requirement, given the serious risks and threats they pose to our foundational constituitonal values and democratic freedoms.

- **Law Society (2019)** Recommendation 5 Lawfulness – "The lawful basis of all algorithmic systems in the criminal justice system must be clear and explicitly declared in advance. "

  "In the course of evidence-taking, Commission became heavily concerned that some systems and databases operating today, such as facial recognition in policing or some uses of mobile device extraction, lack a clear and explicit lawful basis, as well as unclear proven algorithmic performance. This must be urgently examined, publicly clarified and rectified if necessary. While the United Kingdom has more explicit provisions covering algorithmic systems than many other parts of the world, these contains significant omissions and loopholes that need joined-up consideration. Several clarifications and changes to data protection legislation, procurement codes, freedom of information law, equality duties and statutory oversight and scrutiny bodies have been recommended in this

report. These would provide key safeguards to the integrity of criminal justice in the digital age." (Law Commission p 71 concluding remarks)

(4) Vital need for open public discussion, debate and deliberation to decide on what shoudl be regarded as acceptable and unacceptable use, ie to draw some red lines at the national level (principles of ACCOUNTABILITY and OPENNESS)

- Example: no attempt was mde by the MPS prior to live FRT tests to engage the public in any shape or form – only some publicity at a late stage of the trials via press release and website, and sporadic attempts at 'stakeholder engagement' which have been criticised by Liberty and Big Brother Watch as inadequate.

- Nor was there any attempt to engage the broader public, including those resident in the area of test deployment.

(5) Introduce mandatory human rights impact assssment for use of algorithmic systems by the public sector.

- No such requirement at present (although DPIAs mandatory for 'high risk' personal data processing).

- Poor potentially unlawful current practice. Eg MPS live facial recognition tech trial

  "The MPS did prepare a number of impact/risk assessment documents. However, these documents are regarded as inadequate with respect to engagement with human rights law requirements.

  No MPS documents have been seen that clearly set out the justification underpinning the deployment of LFR technology in a manner capable of addressing whether such deployments may be considered 'necessary in a democratic society'.

  Of particular concern is the lack of effective consideration of alternative measures, the absence of clear criteria for inclusion on the watchlist, including with respect to the seriousness of the underlying offence, and the failure to conduct an effective (p 9-10) necessity and proportionality analysis.....Hence MPS' test deployments of LFR technology would not be regarded as 'necessary in a democratic society' if challenged before the courts. "

  No MPS documents have been seen that clearly set out the justification underpinning the deployment of LFR technology in a manner capable of addressing whether such deployments may be considered 'necessary in a democratic society'.

  Of particular concern is the lack of effective consideration of alternative measures, the absence of clear criteria for inclusion on the watchlist, including with respect to the seriousness of the underlying offence, and the failure to conduct an effective necessity and proportionality analysis (p 10)

- Law Society (2019) Sub-Recommendation 4.2 Human Rights by Design – The Government should commission a review into policy options for mandating human rights considerations in technological design within different consequential sectors, including in the criminal justice system. This review should consider how and where human rights impact assessments should be required in public procurement processes.

- Consider mandatory human rights impact assessment for public sector ADM systems. Consider, as a possible model, the Canadian Directive on Automated Decision-Making: new federal legislation requiring public sector to undertake algorithmic impact assessment prior to the use of automated decision-making systems. It requires an algorithmic impact assessment for each ADM System based on assessment criteria that are specified in the Directive from Level – IV. Results of algorithmic impact assessment must be publicly released, and must be updated when functionality or scope of the system changes. Legal requirements vary depending on the assessed impact level (Karen: risk based approach guided by the proportionality principle). For all Canadian ADM Systems used in public sector, REGARDLESS as to their assessed impact, requirements re

  - access, diligence, testing and auditability requirements for software that is licensed;
  - release of any custom source code that is owned by the Government of Canada;
  - testing and monitoring of outcomes by testing data and information used by the ADM System for unintended data biases and other factors that may unfairly impact outcomes before production launch and by developing processes to monitor outcomes of ADM Systems to safeguard against unintentional outcomes and verify compliance with applicable legislation and the Directive itself on a scheduled basis;
  - validating the quality of data collected for and used by the ADM System;
  - security safeguards;
  - legal consultations to ensure that the use of the ADM System complies with applicable laws;
  - providing clients with recourse on decisions of the ADM System so that clients are able to challenge them; and
  - reporting information on effectiveness and efficiency of the ADM System.

KY
30.7.2019