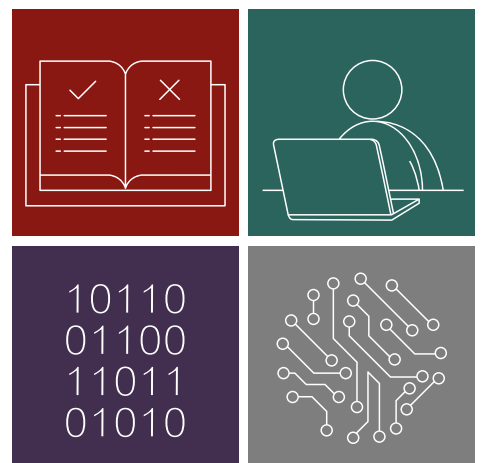


# Artificial Intelligence and Public Standards

## A Review by the Committee on Standards in Public Life

The Committee  
on Standards  
in Public Life





# Artificial Intelligence and Public Standards

A Review by the Committee on  
Standards in Public Life

Chair, Lord Evans of Weardale KCB DL

February 2020



# The Seven Principles of Public Life

The Seven Principles of Public Life (also known as the Nolan Principles) apply to anyone who works as a public office-holder. This includes all those who are elected or appointed to public office, nationally and locally, and all people appointed to work in the Civil Service, local government, the police, courts and probation services, non-departmental public bodies (NDPBs), and in the health, education, social and care services. All public office-holders are both servants of the public and stewards of public resources. The principles also apply to all those in other sectors delivering public services.

## **Selflessness**

Holders of public office should act solely in terms of the public interest.

## **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. They should not act or take decisions in order to gain financial or other material benefits for themselves, their family, or their friends. They must declare and resolve any interests and relationships.

## **Objectivity**

Holders of public office must act and take decisions impartially, fairly and on merit, using the best evidence and without discrimination or bias.

## **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.

## **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.

## **Honesty**

Holders of public office should be truthful.

## **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.



# Chair's Foreword

Dear Prime Minister,

I am pleased to present the 21st report of the Committee on Standards in Public Life on the subject of artificial intelligence and public standards.

Artificial intelligence – and in particular, machine learning – will transform the way public sector organisations make decisions and deliver public services. The government has committed significant resources to this new technology through the AI Sector Deal, which promises to deliver a more accurate, capable and efficient public sector.

Adherence to high public standards will help fully realise the great benefits of AI in public service delivery. By ensuring that AI is subject to appropriate safeguards and regulations, the public can have confidence that new technologies will be used in a way that upholds the Seven Principles of Public Life. Our conclusion from this review is that these principles will remain a valid guide for public sector practice as AI is deployed across all levels of government.

Our recommendations are directed towards three key audiences.

Our message to government is that the UK's regulatory and governance framework for AI in the public sector remains a work in progress and deficiencies are notable. The work of the Office for AI, the Alan Turing Institute, the Centre for Data Ethics and Innovation (CDEI), and the Information Commissioner's Office (ICO) are all commendable. But on the issues of transparency and data bias in particular, there is an urgent need for practical guidance and enforceable regulation.

Regulators must also prepare for the changes AI will bring to public sector practice. We conclude that the UK does not need a specific AI regulator, but all regulators must adapt to the challenges that AI poses to their specific sectors. Government should establish the CDEI as a centre for regulatory assurance to assist regulators in this area.

Upholding public standards will also require action from public bodies using AI to deliver frontline services. All public bodies must comply with the law surrounding data-driven technology and implement clear, risk-based governance for their use of AI.

Artificial intelligence – particularly in the public sector – is the subject of significant media interest and this report will not be the final word on the matter. Nonetheless, we believe our contribution will help the UK public sector uphold public standards as it adopts AI across a wide range of public service delivery.

**Lord Evans of Weardale KCB DL**  
**Chair, Committee on Standards in Public Life**



# Contents

Executive summary	6
List of recommendations	8
Introduction	10
Chapter 1: AI in the UK	12
Chapter 2: AI and the Nolan Principles	16
Chapter 3: Guidance and ethical principles	30
Chapter 4: Regulating AI	39
Chapter 5: The role of public bodies	57
Appendix 1: About the Committee on Standards in Public Life	67
Appendix 2: Terms of reference	68
Appendix 3: Methodology	69



# Executive summary

Artificial intelligence has the potential to revolutionise the delivery of public services, creating an opportunity for more innovative and efficient public service delivery. Machine learning in particular will transform the way decisions are made in areas as diverse as policing, health, welfare, transport, social care, and education.

This review found that the Nolan Principles are strong, relevant, and do not need reformulating for AI. The Committee heard that they are principles of good governance that have stood, and continue to stand, the test of time. All seven principles will remain relevant and valid as AI is increasingly used for public service delivery.

If correctly implemented, AI offers the possibility of improved public standards in some areas. However, AI poses a challenge to three Nolan Principles in particular: openness, accountability, and objectivity. This review examined how public officials and government departments can uphold these principles as AI is increasingly rolled out across our public services.

Our concerns here overlap with key themes from the field of AI ethics. Under the principle of openness, a current lack of information about government use of AI risks undermining transparency. Under the principle of accountability, there are three risks: AI may obscure the chain of organisational accountability; undermine the attribution of responsibility for key decisions made by public officials; and inhibit public officials from providing meaningful explanations for decisions reached by AI. Under the principle of objectivity, the prevalence of data bias risks embedding and amplifying discrimination in everyday public sector practice.

This review found that the government is failing on openness. Public sector organisations are not sufficiently transparent about their use of AI and it is too difficult to find out where machine learning is currently being used in government. It is too early to judge if public sector bodies are successfully upholding accountability. Fears over 'black box' AI, however, may be overstated, and the Committee believes that explainable AI is a realistic goal for the public sector. On objectivity, data bias is an issue of serious concern, and further work is needed on measuring and mitigating the impact of bias.

## Governance and regulation

To uphold public standards, government and public sector organisations should set effective governance to mitigate the risks we have identified. In this sense, AI is a new challenge that can be solved with existing tools and established principles. Public standards can be upheld with a traditional risk management approach.

This is not a challenge that public sector organisations can tackle alone. Government needs to identify and embed authoritative ethical principles and issue accessible guidance on AI governance to those using it in the public sector. Government and regulators must also establish a coherent regulatory framework that sets clear legal boundaries on how AI should be used in the public sector.

Attempts to establish this governance and regulatory framework are emerging and developments are fast-moving. In the area of ethical principles and guidance, the Department for Culture, Media and Sport (DCMS), the Centre for Data Ethics and Innovation (CDEI) and the Office for AI have all published ethical principles for data-driven technology and AI. The Office for AI, the Government Digital Service (GDS), and the Alan Turing Institute have jointly issued A Guide to Using Artificial Intelligence in the Public Sector and draft guidelines on AI procurement. The Information Commissioner's Office (ICO) has also published its Auditing Framework for AI.



In the area of regulation, the use of AI is subject to the provisions of the GDPR, the Equality Act, and sections of administrative law. The government has also established the Centre for Data Ethics and Innovation to advise on regulation.

These developments are positive and are to be welcomed. However, at the time of writing, this review has found that the governance and regulatory framework for AI in the public sector is still a work in progress and one with significant deficiencies.

This is mostly because key documents have only recently been published and government AI institutions are very new. Multiple sets of ethical principles are confusing and the application of each is unclear. Public sector guidance is not yet widely used and public officials with no AI expertise may find it difficult to understand and comply with.

We conclude that a new AI regulator is not needed but existing regulators will need to adapt to face the challenges AI brings. They will need assistance from a central body to do so, but the CDEI does not yet have a clearly defined purpose and is not yet on a statutory footing. Two areas in particular – transparency and data bias – are in need of urgent attention in the form of new regulation and guidance.

### **Our recommendations**

Our recommendations to government and regulators are intended to assist in the development of a stronger and more coherent regulatory and governance framework for AI in the public sector.

We recommend that government should establish consistent and authoritative ethical principles and issue easier to use guidance. Procurement processes should be reformed and the Digital Marketplace should offer greater assistance to public bodies seeking technologies that are compliant with public standards.

Though no new AI regulator is needed, the CDEI should advise regulators on how to adapt to new technologies and be set on an independent statutory footing. The application of anti-discrimination law to AI needs to be clarified and new transparency guidelines are needed. AI impact assessments should be mandatory, published, and set by the CDEI, and new guidelines are needed to enforce transparency.

We also provide recommendations to providers of public services, both public and private, to help them develop effective risk-based governance for AI. During project planning, our recommendations focus on legal and legitimate AI, system design, and diversity. During project implementation, our recommendations cover setting responsibility, internal and external oversight, monitoring and evaluation, appeal and redress, and training and education.

The Nolan Principles remain a valid guide for public sector practice in the age of AI. However, this new technology is a fast-moving field, so government and regulators will need to act swiftly to keep up with the pace of innovation. Our recommendations set out what we believe is needed to ensure the Seven Principles of Public Life are upheld as the public sector transitions into a new AI-enabled age.



# List of recommendations

## Recommendations to government, national bodies and regulators

The Committee makes eight recommendations to government, national bodies and regulators to help create a strong and coherent governance and regulatory framework for AI in the public sector.

### Recommendation 1: Ethical principles and guidance

There are currently three different sets of ethical principles intended to guide the use of AI in the public sector – the FAST SUM Principles, the OECD AI Principles, and the Data Ethics Framework. It is unclear how these work together and public bodies may be uncertain over which principles to follow.

- a. The public needs to understand the high level ethical principles that govern the use of AI in the public sector. The government should identify, endorse and promote these principles and outline the purpose, scope of application and respective standing of each of the three sets currently in use.
- b. The guidance by the Office for AI, the Government Digital Service and the Alan Turing Institute on using AI in the public sector should be made easier to use and understand, and promoted extensively.

### Recommendation 2: Articulating a clear legal basis for AI

All public sector organisations should publish a statement on how their use of AI complies with relevant laws and regulations before they are deployed in public service delivery.

### Recommendation 3: Data bias and anti-discrimination law

The Equality and Human Rights Commission should develop guidance in partnership with both the Alan Turing Institute and the CDEI on how public bodies should best comply with the Equality Act 2010.

### Recommendation 4: Regulatory assurance body

Given the speed of development and implementation of AI, we recommend that there is a regulatory assurance body, which identifies gaps in the regulatory landscape and provides advice to individual regulators and government on the issues associated with AI.

We do not recommend the creation of a specific AI regulator, and recommend that all existing regulators should consider and respond to the regulatory requirements and impact of the growing use of AI in the fields for which they have responsibility.

The Committee endorses the government's intention for CDEI to perform a regulatory assurance role. The government should act swiftly to clarify the overall purpose of CDEI before setting it on an independent statutory footing.

### Recommendation 5: Procurement rules and processes

Government should use its purchasing power in the market to set procurement requirements that ensure that private companies developing AI solutions for the public sector appropriately address public standards.

This should be achieved by ensuring provisions for ethical standards are considered early in the procurement process and explicitly written into tenders and contractual arrangements.





### **Recommendation 6: The Crown Commercial Service's Digital Marketplace**

The Crown Commercial Service should introduce practical tools as part of its new AI framework that help public bodies, and those delivering services to the public, find AI products and services that meet their ethical requirements.

### **Recommendation 7: Impact assessment**

Government should consider how an AI impact assessment requirement could be integrated into existing processes to evaluate the potential effects of AI on public standards. Such assessments should be mandatory and should be published.

### **Recommendation 8: Transparency and disclosure**

Government should establish guidelines for public bodies about the declaration and disclosure of their AI systems.

## **Recommendations to front-line providers, both public and private, of public services**

The Committee makes seven recommendations to front-line providers of public services to help establish effective risk-based governance for the use of AI.

### **Recommendation 9: Evaluating risks to public standards**

Providers of public services, both public and private, should assess the potential impact of a proposed AI system on public standards at project design stage, and ensure that the design of the system mitigates any standards risks identified.

Standards review will need to occur every time a substantial change to the design of an AI system is made.

### **Recommendation 10: Diversity**

Providers of public services, both public and private, must consciously tackle issues of bias and discrimination by ensuring they have taken into account a diverse range of behaviours, backgrounds and points of view. They must take into account the full range of diversity of the population and provide a fair and effective service.

### **Recommendation 11: Upholding responsibility**

Providers of public services, both public and private, should ensure that responsibility for AI systems is clearly allocated and documented, and that operators of AI systems are able to exercise their responsibility in a meaningful way.

### **Recommendation 12: Monitoring and evaluation**

Providers of public services, both public and private, should monitor and evaluate their AI systems to ensure they always operate as intended.

### **Recommendation 13: Establishing oversight**

Providers of public services, both public and private, should set oversight mechanisms that allow for their AI systems to be properly scrutinised.

### **Recommendation 14: Appeal and redress**

Providers of public services, both public and private, must always inform citizens of their right and method of appeal against automated and AI-assisted decisions.

### **Recommendation 15: Training and education**

Providers of public services, both public and private, should ensure their employees working with AI systems undergo continuous training and education.



# Introduction

The Committee on Standards in Public Life (the Committee) was established in 1994. In its first report, the Committee articulated the Seven Principles of Public Life, commonly referred to as the Nolan Principles: selflessness, integrity, objectivity, accountability, openness, honesty and leadership.

The standards landscape has changed significantly since then and the context within which the Committee operates continues to evolve. The Committee, in its 2013 report *Standards Matter*, said:

“The systems and practices of public organisations, the culture and behaviour of public office-holders and the expectations of the public are constantly subject to new influences and constraints, causing them to develop in new and sometimes unexpected ways.”<sup>1</sup>

This is particularly true in the case of new technology. Artificial Intelligence (AI) will fundamentally change the way that government and the public sector operates, and new technology could help design better public policy and deliver more efficient and effective public services.

AI could be used in ways that are uncontroversial. For example, AI could be used to create smart traffic lights where the timing of a red or green light is altered to create the most efficient traffic flow possible.

There is already evidence, however, that AI can be used in more controversial ways. In 2017, a machine learning system was used to see whether there were

any patterns in publicly available data to indicate whether a school was likely to be inadequate or failing.<sup>2</sup> Key correlations were identified and Ofsted began using machine learning to take a more targeted approach to its inspections. Teachers protested and argued that the tool was unfair, lacked transparency, and had the potential to exacerbate pre-existing biases within the education system.

This controversy speaks to wider concerns shared by the Committee. Any change in how the government makes policy decisions and delivers public services must not undermine public standards and the public’s confidence in its institutions. This is particularly important in the context of AI because AI has the potential to change how decisions are made in sensitive policy areas like social care, policing and criminal justice, where the impact on individuals can be significant. It is in these areas that standards will matter most.

The increasing use of AI in public service delivery is also of interest to the Committee because public bodies will not be delivering this change alone. Private companies will often work alongside public bodies to develop and deliver AI solutions. The involvement of commercial organisations in the delivery of public services means that additional care must be given to standards issues.

How the government manages private sector service delivery in a way that exemplifies high ethical standards is not a new issue for the Committee. As the Committee said in its 2018 report *The Continuing Importance of Ethical Standards for Public Service Providers*, the public is right to expect services to be delivered responsibly and ethically, regardless of how they are being delivered, or who is providing those services.<sup>3</sup>

---

1 Committee on Standards in Public Life (2013), *Standards Matter*, Cm 8519, 12. Available at: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/348304/Standards\\_Matter.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/348304/Standards_Matter.pdf)

2 The Behavioural Insights Team (2017), *Using Data Science in Policy*, 14. Available at: [https://www.bi.team/wp-content/uploads/2017/12/BIT\\_DATA-SCIENCE\\_WEB-READY.pdf](https://www.bi.team/wp-content/uploads/2017/12/BIT_DATA-SCIENCE_WEB-READY.pdf)

3 The Committee on Standards in Public life (2018), *The Continuing Importance of Ethical Standards for Public Service Providers*. Available at: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/705884/20180510\\_PSP2\\_Final\\_PDF.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/705884/20180510_PSP2_Final_PDF.pdf)



The Committee undertook this review to ensure that the public sector continues to maintain high standards of conduct as it grapples with the implications and developments made possible by the introduction of AI. This report and its recommendations are designed to help government and public bodies uphold standards as they start to use AI across a wide range of public service provision.

General concerns about privacy in the field of AI are beyond the remit of the Seven Principles of Public Life and are not discussed in this report.

**Chapter 1:** AI in the UK, is an overview of AI, why it matters for the public sector, and current government policy.

**Chapter 2:** AI and the Nolan Principles, considers the relevance of the Nolan Principles in this new age of artificial intelligence and examines the risks and opportunities for openness, accountability and objectivity.

**Chapter 3:** Guidance and ethical principles, examines the recent publication of guidance and ethical principles for using AI in the public sector.

**Chapter 4:** Regulating AI, assesses the legal and regulatory framework for AI in the public sector.

**Chapter 5:** The role of public bodies, outlines how public bodies can manage and mitigate the risk AI poses to public standards through good governance.

The challenge AI poses for standards is real, but it does not require a fundamental reworking of public sector practice. Good, proactive and careful governance is key, and it is the role of government and regulators to encourage this through the development and enforcement of clear and effective AI regulation. This will encourage public institutions to establish suitable governance mechanisms to manage the standards risks associated with the technology they use.

The Committee collected a wide range of evidence for this review, meeting individually with experts in the field from government, academia, and the public and private sectors, holding roundtables, and attending external conferences and workshops. The Committee also held focus groups with members of the public and commissioned public polling on attitudes to AI. The Committee is indebted to all those who contributed to this review.



# Chapter 1: AI in the UK

“Artificial Intelligence is one of the most transformative forces of our time, and is bound to alter the fabric of society.”<sup>4</sup>

**European Commission, Independent High-Level Expert Group on AI**

## 1.1. What is AI?

There is no single uncontested definition of what constitutes AI and the term is used liberally to describe anything from routine data analysis to complex deep neural networks. Whichever definition is used, there is wide agreement that the potential for change and impact on society is immense.

Experts predict it is machine learning that will have the most significant impact and lead to transformative change. Machine learning systems learn from past data by identifying patterns and correlations within it. This allows computers to undertake increasingly complex tasks, like natural language processing and image recognition. These innovations will transform the power of computers to interpret our world. AI systems will be able to analyse and predict human behaviour on an unprecedented scale, in areas as diverse as crime, transport and health.

Complex processes of filtering, cross-referencing and authenticating information, such as the personal data of a benefits claimant to establish their entitlement, could be automated and instantaneous. AI could process and respond coherently to natural language, giving computers the capacity to read legal contracts and converse fluently with human interlocutors. Image recognition could be able to identify distinct people, animals and objects in images and video in real time.

Police forces across the UK are, for example, already using live facial recognition technology to assist in the prevention and detection of crime by identifying wanted criminals.

AI will undoubtedly change the relationship between humans and technology, as well as between citizens and the state. This is because AI allows computers, for the first time, to assist in decision-making processes in a substantive and meaningful way, independent of human judgement.

## 1.2. The scale of AI

The impact of AI across the public and private sectors is potentially vast. These advances in computing capability will revolutionise areas such as finance, energy, health, education and agriculture. The Office for AI (see opposite) estimates that AI could add £232 billion to the UK’s economy by 2030, boosting productivity in some industries by 30%.<sup>5</sup> AI is also an international issue. Over 25 countries have published an AI strategy, and the European Union, United Nations, and OECD have all taken a close interest in AI governance and ethics. The question of how AI can be used effectively and ethically is of global concern and there would be benefit to the UK working with its international partners in a shared approach.

Government has a particular responsibility to exercise care around the use of AI in the public sector. Citizens can choose not to use a particular private company’s products or services, but citizens cannot always opt out of public service delivery. Public sector AI will be funded by taxpayers’ money, and in some cases AI will be part of the operation of the law. The use of AI in the UK public sector must follow the Seven Principles of Public Life, which outline the ethical values to which the public sector should adhere.

4 Council of Europe (2018), Draft Recommendation of the Committee of Ministers to member States on human rights impacts of algorithmic systems. Available at: <https://rm.coe.int/draft-recommendation-on-human-rights-impacts-of-algorithmic-systems/16808ef256>

5 Office for Artificial Intelligence (AI) (2019), AI Sector Deal One Year On. Available at: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/819331/AI\\_Sector\\_Deal\\_One\\_Year\\_On\\_Web\\_.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/819331/AI_Sector_Deal_One_Year_On_Web_.pdf)



### 1.3. UK policy on AI

In 2017, the government published an independent review led by Professor Dame Wendy Hall and Jerome Pesenti (the Hall-Pesenti review) on how the AI industry could be developed in the UK.<sup>6</sup> The review made a number of important recommendations to improve access to skills and data, maximise AI research and support the uptake of AI.

Following the Hall-Pesenti review, the government published its 2017 Industrial Strategy,<sup>7</sup> which identified AI and data as one of four ‘Grand Challenges’ to modernise the UK economy.<sup>8</sup> The AI Sector Deal was published in 2018.<sup>9</sup> It made clear that the government sees AI as a “huge global opportunity” and wants to become a global leader in AI and data-driven technology.

The AI Sector Deal led to the creation of three new institutions: a government Office for AI; an industry-led AI Council; and the Centre for Data Ethics and Innovation (CDEI).

#### The Office for AI

The Office for AI is a joint office sponsored by the Department for Digital, Culture, Media and Sport (DCMS), and the Department for Business, Energy and Industrial Strategy (BEIS).

Their role is to oversee the implementation of the AI Sector Deal, which is part of the AI and Data Grand Challenge. They have recently published the AI Guide on implementing AI in the public sector, as well as draft guidelines for AI procurement. The Office for AI aims to increase adoption of AI across the private and public sectors.

#### The AI Council

The Hall-Pesenti review recommended that government should work with industry to establish an industry-led AI Council to advise government on AI.

The AI Council is an independent expert committee that advises government on how to promote the growth of AI in the UK. It includes representatives from the public and private sectors.

#### The Centre for Data Ethics and Innovation (CDEI)

CDEI is an independent public sector body established by DCMS to advise government on artificial intelligence and other data-driven technologies. It is tasked to help develop the right regulation and governance for data-driven technology.

CDEI is currently undertaking thematic reviews on issues of public concern, such as data bias and online targeting. Their remit includes both the public and private sectors.

CDEI is not yet on a statutory footing and its final status is yet to be determined. The government has said it intends to place CDEI on a statutory footing after its initial phase of operation.

- 
- 6 Professor Dame Wendy Hall and Jerome Pesenti (2017), Growing the Artificial Intelligence Industry in the UK. Available at: <https://www.gov.uk/government/publications/growing-the-artificial-intelligence-industry-in-the-uk>
- 7 Department for Business, Energy and Industrial Strategy (2017), Industrial Strategy: Building a Britain fit for the Future, White Paper. Available at: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/664563/industrial-strategy-white-paper-web-ready-version.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/664563/industrial-strategy-white-paper-web-ready-version.pdf)
- 8 Department for Business, Energy and Industrial Strategy (2019), The Grand Challenges, Policy Paper. Available at: <https://www.gov.uk/government/publications/industrial-strategy-the-grand-challenges/industrial-strategy-the-grand-challenges#artificial-intelligence-and-data>
- 9 Department for Business, Energy and Industrial Strategy (2018), Industrial Strategy: Artificial Intelligence Sector Deal. Available at: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/702810/180425\\_BEIS\\_AI\\_Sector\\_Deal\\_4\\_.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/702810/180425_BEIS_AI_Sector_Deal_4_.pdf)



These new institutions will work alongside other public bodies created before the Sector Deal. Notably, the Alan Turing Institute was made the national institute for artificial intelligence and data science by the government in 2017, in response to a recommendation made in the Hall-Pesenti review. Currently based at the British Library, the Institute convenes academics and industry to research AI and its impact on society. The Institute has been working with the Office for AI to produce guidance on how to use AI ethically and safely.

The Government Digital Service (GDS) also has a significant role in shaping AI policy. GDS is part of the Cabinet Office and is responsible for digital transformation across government. GDS currently sets and enforces standards for digital technology, including around procurement.

In 2019, GDS published joint guidance with the Office for AI and the Alan Turing Institute on how to use AI in the public sector.<sup>10</sup>

AI used in the UK public sector is also subject to the Data Ethics Framework, published by the Department for Digital, Culture, Media and Sport (DCMS) in 2018.<sup>11</sup> The framework guides appropriate data use in government and the wider public sector, and is aimed at anyone working directly or indirectly with data, including data scientists, policymakers and operational staff. It is not binding but builds on the core values of the Civil Service Code – integrity, honesty, objectivity and impartiality – to encourage ethical data use, build better services and inform policy.

DCMS has also published guidance around each principle, and a Data Ethics Workbook to help practitioners align their work with the framework's principles.

### The Data Ethics Framework principles

1. Start with clear user need and public benefit
2. Be aware of relevant legislation and codes of practice
3. Use data that is proportionate to the user need
4. Understand the limitations of the data
5. Ensure robust practices and work within your skillset
6. Make your work transparent and be accountable
7. Embed data use responsibly.<sup>12</sup>

10 Office for AI and Government Digital Service (GDS) (2019), A guide to using AI in the public sector. Available at: <https://www.gov.uk/government/collections/a-guide-to-using-artificial-intelligence-in-the-public-sector>

11 Department for Digital, Culture, Media and Sport (2018), Data Ethics Framework. Available at: <https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework>

12 Same source



#### 1.4. Where is AI being used in the UK public sector?

Despite generating much interest and commentary, our evidence shows that the adoption of AI in the UK public sector remains limited. Most examples the Committee saw of AI in the public sector were still under development or at a proof-of-concept stage.

The Committee heard, however, that many public bodies are beginning to look at how they can use AI to deliver better public services. Our evidence showed that healthcare and policing currently have the most developed AI programmes, with technology being used, for example, to identify eye disease and to predict reoffending rates, though levels of system maturity differ across NHS trusts and police forces.

The Committee found that the Judiciary, the Department for Transport (DfT) and the Home Office are examining how AI can increase efficiency in service delivery. The Home Office also told us that they are currently looking at the governance structures that need to be in place when AI is used in the public sector.

The Committee was told that local government is currently innovating with AI systems in education, welfare and social care. Hampshire County Council, for example, is trialling the use of smart devices, such as Amazon Echo, in the homes of adults receiving social care, to bridge the gap between visits from professional carers.<sup>13</sup> The Guardian reported that one-third of councils use algorithmic systems to make welfare decisions.<sup>14</sup>

It is the view of the Committee, however, that obstacles to widespread and successful adoption remain significant. Public policy experts frequently told this review that access to the right quantity of clean, good-quality data is limited, and that trial systems are not yet ready to be put into operation. It is our impression that many public bodies are still focusing on early-stage digitalisation of services, rather than more ambitious AI projects.

Multiple contributors to the review also commented that the lack of a clear standards framework – including in law and regulation – meant that organisations did not have the confidence to use AI. While standards and regulation are often seen as barriers to innovation, the Committee believes that implementing clear ethical standards around AI may accelerate rather than delay adoption, by building trust in new technologies among public officials and service users.

13 SA Mathieson (2019), 'I feel in control of my life: Alexa's new role in public service', The Guardian. Available at: <https://www.theguardian.com/society/2019/feb/07/control-life-alexa-role-public-service-chatbots-councils>

14 Sarah Marsh (2019), 'One in three councils using algorithms to make welfare decisions', The Guardian. Available at: <https://www.theguardian.com/society/2019/oct/15/councils-using-algorithms-make-welfare-decisions-benefits>



# Chapter 2: AI and the Nolan Principles

## 2.1. The Seven Principles of Public Life

The Seven Principles of Public Life apply to anyone who works as a public office-holder. They also apply to those in the private sector delivering public services. These well-established principles set the standards across the whole of public service.

The Nolan Principles have been widely accepted as the basis of good practice throughout the public sector. They are mentioned explicitly in the UK, Scottish, Welsh and Northern Irish Ministerial Codes, are included in the corporate documentation of a large number of public sector organisations, and form the basis of the codes of conduct required of all local authorities. Some organisations, like the Civil Service, have adapted the principles to their own particular context.<sup>15</sup>

The way public bodies view ethical standards has, however, changed over the years. The Committee noted in its 2013 report *Standards Matter* that in many organisations, the debate on ethical standards has shifted from an emphasis on personal standards to an approach which places greater weight on managing risks to standards in an organisation as a whole.<sup>16</sup>

The increased adoption of AI will bring new challenges to the practices of public organisations and the behaviour of public office-holders, as well as affecting the expectations of the public. As part of this review, the Committee examined whether artificial intelligence would require a fundamental rethinking of public standards.

There was a general consensus among contributors to this review that the Nolan Principles are strong, relevant, and do not need reformulating for AI systems. The Committee heard that they are

principles of good governance that have stood, and continue to stand, the test of time.

This was partly because they are already well known and embedded within the cultures of organisations across the public service, and also because they are highly relevant in terms of the ethical challenges AI will have to meet.

The Committee is aware that while principles are important, they are not sufficient on their own as a complete guide for behaviour in public life. To ensure that ethical principles generate changes in behaviour, they need to be elaborated in codes of conduct and guidance and implemented through policy and governance. The application of the principles in AI may not be self-evident, and in some cases it will be unclear exactly how public officials should uphold these ethical principles in practice.

The Committee heard that more needs to be done to achieve this behavioural change. Codes and principles should be embedded into current practices through better governance, leadership and education. The Committee also heard that internal systems for upholding standards in public bodies should be supported by independent scrutiny.

The following chapters of this report outline what government and public bodies can do to translate the Seven Principles into practice for the use of AI.

## 2.2. Where is AI likely to affect public standards?

All of the Seven Principles of Public Life must be upheld when using AI in the public sector. Three principles are particularly relevant: openness, accountability and objectivity.

15 Civil Service Code of Conduct (2015). Available at: <https://www.gov.uk/government/publications/civil-service-code/the-civil-service-code>

16 Committee on Standards in Public Life (2013), *Standards Matter*, Cm 8519. Available at: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/348304/Standards\\_Matter.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/348304/Standards_Matter.pdf)



## 2.3. 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.**

### 2.3.1. Why openness matters

The public can only scrutinise and understand the decisions of government and public bodies if they have access to information about the evidence, assumptions and principles on which policy decisions have been made. Citizens should have access to information about government policies that affect their lives.

“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 rulemaking not subject to appropriate procedural safeguards or societal oversight.”<sup>17</sup>

**Law Society Report, Algorithms in the Criminal Justice System**

Access to this information also facilitates fair and informed public debate. Democratic society cannot make meaningful decisions if the government and the wider public sector are not open about their processes, capabilities and functions.

The use of artificial intelligence in the public sector may also change the relationship between citizens and the state. Surveillance technologies like facial recognition could increase the power of the state and other actors to monitor citizens’ lives. Machine learning systems are also likely to shift the impetus of public service delivery from reaction and redress to prediction and prevention. These are not just questions of policy. These issues raise fundamental questions about democracy and human rights. The Committee considers that government openness about its use of AI is essential.

“States should engage in inclusive, interdisciplinary, informed and public debates to define what areas of public services profoundly affecting access to or exercise of human rights may not be appropriately determined, decided or optimised through algorithmic systems.”<sup>18</sup>

**The Council of Europe’s draft Guidelines for States on actions to be taken vis-à-vis the human rights impacts of algorithmic systems**

While members of the public expressed a clear preference for openness in the focus groups held for this review, they also understood the need to judge the particular context in which AI is being used. Too much information was seen as being as unhelpful as too little. Openness does not mean that every detail around every use of AI in the public sector must be made public, and it may not be necessary or desirable to publish the source code for an AI system. Nonetheless, it is the view of the Committee that fundamental information about the purpose of the technology, how it is being used, and how it affects the lives of citizens must be disclosed to the public.

17 The Law Society (2019), Algorithms in the Criminal Justice System. Available at: <https://www.lawsociety.org.uk/support-services/research-trends/documents/algorithm-use-in-the-criminal-justice-system-report/>

18 Council of Europe (2018), Draft Recommendation of the Committee of Ministers to member States on human rights impacts of algorithmic systems’. Available at: <https://rm.coe.int/draft-recommendation-on-human-rights-impacts-of-algorithmic-systems/16808ef256>



### 2.3.2. How open is government about its use of AI?

Evidence submitted to this review suggests that at present the government and public bodies are not sufficiently transparent about their use of AI. Many contributors, including a number of academics, civil society groups and public officials said that it was too difficult to find out where the government is currently using AI. Even those working closely with the UK government on the development of AI policy, including staff at the Alan Turing Institute and the Centre for Data Ethics and Innovation, expressed frustration at their inability to find out which government departments were using these systems and how.

“We are not aware of any body with systematic knowledge of where automated decision-making tools are being used in the public sector.”

**Centre for Data Ethics and Innovation**

The government does not publish any centralised audit identifying and making publicly available the extent of AI use across central government or the wider public sector. Public bodies rarely take a proactive approach to publishing information about their AI systems. Most of what we know is the result of work by journalists and academics who often have to rely on Freedom of Information Requests (FOIs), parliamentary questions or poorly formatted procurement data.<sup>19</sup>

Our evidence suggests that this lack of transparency is particularly pressing in policing and criminal justice. Many contributors said that they had trouble accessing important information about the use of new technologies in this area. This is particularly concerning given that surveillance technologies like automated facial recognition have the potential to undermine human rights.

“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.”

**Professor Karen Yeung, Interdisciplinary Professorial Fellow in Law, Ethics and Informatics, University of Birmingham Law School and School of Computer Science**

Transparency is further complicated by the use of private sector commercial organisations in the development and provision of AI systems for use in the public sector, particularly as private companies may cite the need for commercial confidentiality to avoid certain forms of disclosure. This is concerning given that skills and resource constraints mean that public bodies are more likely to contract private companies to deliver AI services than to develop them in-house. In such cases, as elsewhere in public-private partnerships, the Principles of Public Life are binding on all those who provide services financed by public money.

19 Bureau of Investigative Journalism (2019), Government Data Systems: The Bureau Investigates. Available at: <https://www.thebureauinvestigates.com/stories/2019-05-08/algorithms-government-it-systems>



“Transparency – and therefore accountability – over the way in which public money is spent remains a very grey area in the UK...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.”<sup>20</sup>

**Dr Crofton Black, Government Data Systems: The Bureau Investigates, The Bureau of Investigative Journalism**

This lack of transparency poses a clear risk to standards. If public bodies are not sufficiently open about where and how they are using these systems, the public will not be able to scrutinise and hold accountable those institutions which use AI in dubious or controversial ways. Without a well-informed public debate around AI, it is also likely that the UK will lack a common consensus on where and how the technology can be used for good in the public sector.

“Much of the public simply don’t yet know enough about how AI or automation works, or where innovations might be used, to make an informed decision on whether they support or oppose them. This creates a vacuum of information, into which negative narratives about Britain’s future are just as likely to take root as positive ones.”<sup>21</sup>

**Mark Kleinman, Professor of Public Policy and Director of Analysis at the Policy Institute, King’s College London**

## 2.4. Accountability

***Holders of public office are accountable to the public for their decisions and must submit themselves to the scrutiny necessary to ensure this.***

### 2.4.1. How can organisations be accountable when using AI?

Accountability is about holding individuals and organisations responsible for how any AI application is being used in the public sector. Yet applying the principle in this context is complicated by the fact that the outcome of an AI system will not simply be the product of the software itself, or any single decision-maker.

This is because the success or failure of an AI system may be the product of one or several components. In most cases, a system failure will be the result of multiple factors, and responsibility will not be easily apportioned.

All public officials responsible for part of an AI-assisted process will have a degree of professional accountability for their areas of responsibility. However, ultimately, accountability for AI systems lies with senior leadership who oversee AI projects and set governance for its effective and ethical use. It is senior leadership who should be held accountable if their staff are not sufficiently trained, if they have not implemented proper checks on the quality of data, or if they have approved the deployment of a system that prevents public officials from altering the basis on which AI makes a decision. Rather than lying with any single designer, system-builder, or operator of an AI system, accountability has to rest with those who choose to adopt and implement the system as part of their responsibility for public service delivery. It is the role of senior leadership to ensure that suitable governance is in place for any risks a system poses.

20 Same source

21 ‘People not robots are the key to the fourth industrial revolution’, Mark Kleinman, accessed at: <https://www.kcl.ac.uk/news/people-not-robots-are-the-key-to-the-fourth-industrial-revolution>



Senior leadership should then be held accountable for decisions an AI system takes.

#### 2.4.2. Should AI have full responsibility for decision-making in the public sector?

The issue of responsibility concerns how much human involvement there is in each individual decision taken by an AI system. As machine learning systems will be able to make decisions in many areas of public service delivery without any human involvement, the rise of AI means public bodies will need to assess the extent to which public officials should be involved in a decision-making process. It also raises the question of who or what is ultimately responsible for the outcomes of AI-enabled decisions. Both questions are key for accountability. Public office-holders and the public will need to understand how decisions are made and on the basis of what evidence.

It may become increasingly difficult to assign human responsibility to an automated decision-making process if AI can make decisions autonomously and automatically. In the future AI could be granted legal personhood and be held liable for its own decisions in the same way that a private company is. This would require a radical reworking of the law.

“When you have a non-human decision-maker, can you always ascribe the outcome to a human? If you cannot then you have a gap where there is no legal liability. One could stretch existing laws around negligence and vicarious liability, but the more independently AI takes decisions, the harder it will be to tie decisions back to human beings.”

**Jacob Turner, Barrister and Author of Robot Rules: Regulating Artificial Intelligence**

Most experts consulted for this review rejected legal personhood for AI. Instead, policymakers, technologists and ethicists all told the Committee that retaining an element of human responsibility was a prerequisite for upholding high ethical standards in AI-enabled public services. This is a common theme in current AI ethics codes, which usually make explicit that AI should be human-centric, uphold human agency and respect human autonomy.

In polling undertaken for this review, there was also a clear preference for upholding a degree of human responsibility in automated decision-making in the public sector. 69% of those polled said that they would be more comfortable with a public body using AI if a human was using their professional expertise to make a final judgement on any decision. Participants in focus groups also took this view and said that the lack of human involvement in a decision-making process would be unnerving.<sup>22</sup>

Retaining a degree of human involvement and responsibility for automated decision-making is also likely to help uphold public standards in practice. There will be more of an incentive for public officials to monitor and check their AI systems if an official has to answer to the public for the outcome of an automated decision. The Committee therefore believes that public officials should be in control of AI, retain some involvement in all automated decision-making processes, and take responsibility for decisions made by AI systems.

The extent and nature of this responsibility will vary. The Committee heard about a number of models for upholding human responsibility in automated decision-making. Some contributors told the Committee that forcing public officials to directly intervene in all simple, automated decisions was neither fair nor plausible, particularly where intervention would be unnecessary or obstructive.



More limited forms of oversight, such as monitoring and evaluation, would likely still be necessary in this context, and would allow public officials to identify and remedy potential problems within the system. This was seen as a fairly limited form of oversight, as an automated decision could still occur without significant human involvement.

The Committee heard from some experts that ‘human-in-the-loop’ models are helpful for retaining a degree of human control over an automated decision-making system. In a ‘human-in-the-loop’ system, a public official can intervene in the decision-making process of a machine. This means that AI works more or less autonomously, but that a human can observe how different variables are weighted and intervene where necessary. In these systems, humans are likely to be involved in the training process of the algorithm, continuously testing and tuning the data in order to achieve better results.

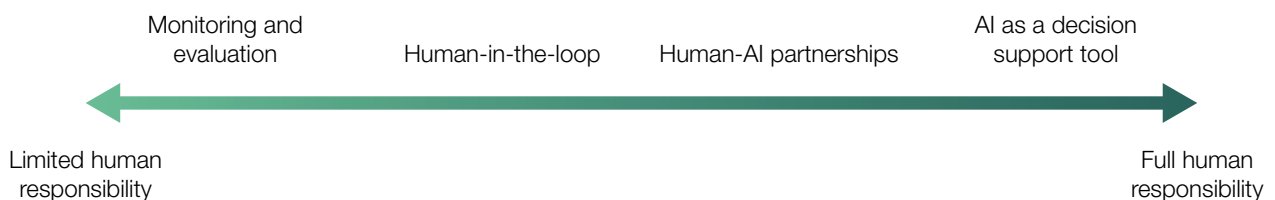
Others rejected the concept of ‘human-in-the-loop’ as too AI-centric. Some contributors argued that these models portrayed AI as immutable, because they ask individuals to shoehorn human judgement into machine learning systems so that human responsibility is protected. This was seen as inverting a ‘human-centric’ approach. Contributors

spoke instead of a more interactive partnership between human and machine, where the outcome of an automated decision is the equal product of human and AI involvement.<sup>23</sup> This was seen as useful for eradicating the potential flaws of both human and AI decision-making processes, and for enhancing the quality and accuracy of the decision as a whole.

“Rather than focusing on the concept of humans-in-the-loop, we need to think carefully about the end-to-end process and ensure that we think about how AI and humans work together to deliver efficiencies and better results.”

**Sana Khareghani, Head, Office for AI**

Those who saw the relationship between human and machine in this way suggested that there should be an element of human control at every stage of the AI process, from design, through procurement, to the deployment of an AI system. Contributors suggested that this whole-systems approach would help mitigate the risk of an accountability gap, where it is unclear which public officials, if any, are responsible for an automated decision.



23 N. R. Jennings, L. Moreau, D. Nicholson, S. Ramchurn, S. Roberts, T. Rodden and A. Rogers (2014) “On human-agent collectives” Communications of ACM 57 (12) 80-88



“If you are saying that there may be some decisions that need to be made so rapidly that the machine makes the decision (if it has been appropriately codified), there is still human accountability at the design stage and in the verification and validation of the AI system before it is put into use. This means you may not have an accountability gap as ultimately a human is still accountable at the design and testing stages.”

**Fiona Butcher, Fellow, Defence Science and Technology Laboratory, Ministry of Defence**

Many contributors took the view that AI should not retain any role in making a final decision, particularly where the adverse effects on an individual could be significant. They suggested instead that AI should be thought of as a decision-support tool, rather than a decision-making system. For example, an AI system that identifies a malignant melanoma should not be seen as making a decision in conjunction with a medical professional, but as making a recommendation to a doctor who retains final discretion on diagnosis and treatment.

Models for upholding human responsibility can be placed on a spectrum, from humans having limited to full responsibility for an automated decision, as shown below. Senior leadership will have to choose which level of responsibility is most appropriate for the application of AI in their organisation (see chapter 5).

### 2.4.3. Should public bodies provide explanations for AI decisions?

Automated decision-making systems usually work by analysing large quantities of data to find patterns and correlations between variables and outcomes. These patterns are used to generate insight, which is used to inform a recommendation or a decision. While the outcome of an AI system will be clear, the process by which it comes to a decision will often not be. The Committee heard that some more complex forms of machine learning cannot show how they determine which variable caused which outcome, or in some cases, what those variables are. AI systems that are opaque in this way are often referred to as ‘black boxes’.

“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.”<sup>24</sup>

**Leverhulme Centre for the Future of Intelligence, University of Cambridge**

The term ‘explainability’ is typically used to describe the extent to which an AI system’s decision-making process can be understood. The levels of explainability necessary will vary across systems, but to uphold accountability, an AI system will need to provide some kind of explanation of its decision-making process.

It was the view of most contributors to this review that the black box problem is largely an avoidable issue. The Committee heard that most machine learning systems deployed in the public sector will be processing data held in simple, readable formats, such as demographic data. Here, a less complex, more explainable system could be used as it would deliver results of a similar level of accuracy to more complex, unexplainable systems.

24 Written evidence 9 (Leverhulme Centre for the Future of Intelligence, University of Cambridge)



“If you stick with a simpler model which is inherently interpretable, you are not going to sacrifice that much on accuracy but you are going to keep the benefits of understanding the variables you are using and understanding how the model works.”

**Dr Reuben Binns, Postdoctoral Research Fellow in AI, ICO**

Very complex systems, such as those based on neural networks, can make it hard to follow the logic of the system. However, in such cases, there may be a trade-off between accuracy and explainability. The Committee heard that these technologies are unlikely to be used in the near future in the public sector. Where these systems are used, public officials will need to be able to justify why their need for such complex systems outweighs the requirement for transparency.

Many contributors saw public officials and private companies choosing not to provide an explanation as a greater obstacle than technical capability. The most significant risk for public standards is that officials and companies will fail to include provisions for explainability when designing their systems even though it would be technically possible to do so.

“I think something we need to be challenging ourselves on is whether the lack of transparency and the lack of explainability is a real necessity for the system or whether it is bad design...sometimes there is a challenge to be made of vendors and people who are building the system.”

**Simon McDougall, Executive Director, Technology Policy and Innovation, ICO**

Our evidence showed this could be for multiple reasons. Building in provisions for explainability could increase the cost of the system.

The Committee also heard that private providers of public services may not want to reveal the intricacies of their systems in order to protect their intellectual property rights and commercial secrets. This ‘commercial black box’ was cited by some as a greater obstacle to transparency than technical opacity.

Private companies developing AI software consulted for this review told the Committee that they often had the capability to make their products and services more explainable, but that they were rarely asked to do so by those procuring technology for the public sector. The Committee was told that requirements for technical transparency are not usually included in procurement tenders and contracts.

“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.”<sup>25</sup>

**Centre for Data Ethics and Innovation**

Overall, the evidence submitted to this review suggests that technical obstacles to public bodies providing explanations for AI-enabled decisions are currently small. It should be possible for citizens to obtain meaningful explanations in policy areas as diverse as healthcare, policing and social care. To achieve this, explainability needs to be considered in the early stages of project development and design, and during procurement processes, by those commissioning the technology for use in the public sector.



If explanations are provided, AI could present an opportunity to enhance accountability and openness in public services. Understanding the reasons behind human decision-making is often fraught with difficulty. It will not always be possible to understand, for example, how a public official came to their decision about a benefits claimant, or why their judgement was correct. When AI is used alongside human judgement, it may help provide greater clarity over which variables informed a decision.

Given that public bodies will, more often than not, be able to provide explanations for AI decisions, the key question is how, when and to whom explanations should be provided. Public bodies should note that the provision of an explanation appeals to the general public. In polling done for this review, 51% of those polled said that the provision of “an easy-to-understand explanation for the AI software’s decision” would make them “much more comfortable” or “a bit more comfortable” with using AI in the public sector.<sup>26</sup>

“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...consideration should be given to the circumstances in which reasons for an explanation of the output may be required.”<sup>27</sup>

**Marion Oswald, Senior Fellow in Law and Director of the Centre for Information Rights, University of Winchester**

Overall, the Committee heard that the type of explanation necessary, or indeed whether an explanation was needed at all, was dependent on context. High impact decisions, such as those that have the potential to affect a citizen’s rights or grant access to a service, are more likely to require clear explanations that give an account of the rationale, reasons and individual circumstances for a specific automated decision. Low-impact decisions, such as those made to increase administrative efficiency, are less likely to require an explanation beyond a statement that outlines the general functionality of an automated system. The ICO’s Project ExplA/n report also found that in some contexts, for example in healthcare, accuracy was seen as more important than explainability.<sup>28</sup>

The Committee heard that there were also valid reasons not to disclose how an AI system came to a decision. There were concerns about individuals being able to manipulate systems for desired outcomes if public bodies were too transparent about what variables were used to inform a decision. A regulator, for example, may not want to provide an explanation for an AI system used to identify non-compliance with the law in case companies learn how to evade detection of non-compliance.

These concerns may well be valid reasons not to provide explanations to service users in certain contexts. However, those reasons must be shown to be legitimate and not an excuse to implement unexplainable systems. The burden of proof should always be on a public official to justify in clear terms why the benefits of explainability are outweighed by the possible detriment disclosure could cause. In such cases, this follows the principle of openness.

26 Appendix 3, 81

27 Written evidence 4 (Marion Oswald)

28 The Information Commissioner’s Office (ICO) (2019), Project Explain, Interim report. Available at: <https://ico.org.uk/media/about-the-ico/documents/2615039/project-explain-20190603.pdf>





## 2.5. Objectivity

***Holders of public office must act and take decisions impartially, fairly and on merit, using the best evidence and without discrimination and bias.***

### 2.5.1. Data bias

Data can be collected on multiple aspects of our world, from the speed of a car to somebody's personal preferences. Millions of data points together can reflect more complex scenarios, such as city-wide traffic jams or the voting habits of demographic groups.

But data may not always be representative. It is well understood that AI has the potential to produce discriminatory effects if a data set is in some way flawed or an algorithm operates in a biased way. Machine learning identifies patterns in past data and makes current decisions based on those patterns, so AI systems have the potential to entrench or amplify historic biases. AI systems could exacerbate biases against protected characteristics, such as race or sex, and make discriminatory outcomes against those characteristics more likely.

Imagine a machine learning system deployed by a company to filter job applications by scanning individuals' CVs. The system has 'learned' what makes a successful applicant by processing the CV data of past successful applicants – when the recruitment process was run by humans – to determine what each successful applicant has in common. In theory, the system should identify educational qualifications, relevant experience, and seniority in previous roles as key criteria. It should then use these criteria to filter new applications.

But this theory only holds if educational qualifications, relevant experience and seniority were the determinants of successful applications when humans ran the company's recruitment process. If those humans were biased themselves – say, for example, they favoured male applicants over female applicants – the machine learning system would inevitably replicate that bias. This was the case with an Amazon machine learning system developed in 2014, “which effectively taught itself that male candidates were preferable”.<sup>29</sup>

Sampling errors can also produce discriminatory outcomes. A machine learning tool designed to diagnose skin cancer that has only been trained on white skin will be less accurate on other skin colours. This bias in the training data may not be the result of active human prejudice, but it will result in a discriminatory outcome: the system is more likely to misdiagnose BAME people.

“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.”<sup>30</sup>

**British Computer Society**

29 Maya Oppenheim (2018), ‘Amazon scraps “sexist AI” recruitment tool’, Independent. Available at: <https://www.independent.co.uk/life-style/gadgets-and-tech/amazon-ai-sexist-recruitment-tool-algorithm-aw8579161.html>

30 Written evidence 11 (British Computer Society)



Ultimately, AI systems are only as good as the data we put into them. ‘Bad’ data can contain implicit and explicit racial, gender or ideological biases. When this data is used to train machine learning algorithms, these biases find their way into the AI systems we design, which can result in discriminatory decisions. The Committee heard that machine learning systems could potentially discriminate on the basis of any variable it identifies when processing data.

“Decision-making, algorithmic or otherwise, can of course also be biased against characteristics which may not be protected in law, but which may be considered unfair, such as socio-economic background. In addition, the use of algorithms increases the chances of discrimination against characteristics that are not obvious or visible. For example, an algorithm might be effective at identifying people who lack financial literacy and use this to set interest rates or repayment terms.”<sup>31</sup>

**Centre for Data Ethics and Innovation,  
Interim Report on Data Bias**

### 2.5.2. How will data bias affect objectivity?

The principle of objectivity requires government and public bodies to act and take decisions impartially, without discrimination or bias. Data bias therefore poses a direct risk to public standards. The introduction of automated machine learning systems in areas such as policing, immigration and healthcare risks inadvertently introducing or amplifying discrimination in sensitive policy areas. From a standards perspective, and in the eyes of the law, discrimination via algorithm is no less of an offence than discrimination by a public official.

“The statistics speak for themselves. We know that you are eight times more likely to be subject to stop and search in the UK if you are black. If you are building an algorithm on these statistics, that is a huge problem.”

**Sandra Wachter, Associate Professor  
and Senior Research Fellow,  
Oxford Internet Institute**

Civil liberties groups fear discriminatory machine learning systems are already in use in the UK. In June 2019, the Financial Times reported that the Home Office used an algorithmic tool to stream visa applications. The potential for the streaming tool to replicate historic bias is clear: if officials had previously discriminated against applicants from certain countries, the streaming tool would do so too.<sup>32</sup>

Data bias is a well-known phenomenon that frequently features in the media. In focus groups undertaken for this review, there was evidence that the public are aware of the issue. When given the example of predictive policing software, participants immediately mentioned the risk of biased profiling, despite assumptions that computers are inherently neutral or objective. For most participants data bias was seen as a more significant issue than a lack of explainability or human responsibility. Of the three issues, data bias appeared to have the greatest potential to delegitimise the use of AI in the public sector in the eyes of the general public.

However, policy experts often qualified negative perceptions of data bias with three considerations. They made the point that data bias can be used to measure and reveal discrimination in existing public sector practices. Often, marginalised groups complain of systematic discrimination from public

31 Centre for Data Ethics and Innovation (CDEI) (2019), Interim Report: Review into bias in algorithmic decision-making. Available at: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/819168/Interim\\_report\\_-\\_review\\_into\\_algorithmic\\_bias.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/819168/Interim_report_-_review_into_algorithmic_bias.pdf)

32 Helen Warrell (2019), ‘Home Office under fire for using secretive visa algorithm’, Financial Times. Available at: <https://www.ft.com/content/0206dd56-87b0-11e9-a028-86cea8523dc2>



bodies, but lack the statistical tools to measure bias. Machine learning tools will probably be able to reveal that discrimination, and potentially expose new, unknown biases.

“Some of our existing systems are designed in a way that makes it impossible to measure bias...One of the good things about machine learning technologies is that they have exposed some bias which has always been there.”

**Professor Helen Margetts, Professor of Society and the Internet at the University of Oxford and Director of the Public Policy Programme, The Alan Turing Institute**

Second, it was frequently mentioned that AI systems will be no more biased than the human processes they are replacing. For some, this meant criticism of the discriminatory impact of AI systems was overblown, as AI systems were not likely to be significantly worse than what already exists.

“Right now we are more likely to be replacing a human process with an AI process. All us humans are bringing a whole suitcase of preconceptions, prejudices and baggage along with us to that decision, some conscious and some unconscious. As we talk around bias in AI – and there is plenty of stuff to talk about – we have to keep in mind we are not moving from a beautiful neutral model.”

**Simon McDougall, Executive Director, Technology Policy and Innovation, ICO**

Third, the Committee heard that if technologists can successfully identify and minimise bias, AI has the potential to be more objective than humans currently are. Contributors cited a famous ‘hungry judge’ study, which found that judges were more likely to issue harsher decisions just before lunch.<sup>33</sup> AI systems on the other hand do not get hungry. Though confidence in the possibility of eradicating data bias was mixed, some contributors said they could foresee a future where, in some areas, a duty of objectivity could require public bodies to use AI systems rather than humans.

“I think we have to start from the point of view that we are dealing with biased systems usually anyway. It is one of the hopes of artificial intelligence that it might be able to reduce bias in certain areas and, certainly, provide lots more ways of systematically thinking about measuring that bias.”

**Dr Jonathan Bright, Senior Research Fellow, Oxford Internet Institute**

### 2.5.3. Mitigating and managing data bias

AI experts suggested a range of methods to manage data bias. Chief among these was the need to ensure diversity in AI teams. A workforce composed of a single demographic is less likely to check for and notice discrimination than diverse teams. At every stage – from the design of a product to its deployment – diversity was seen as a necessity. The Committee heard that while data bias may create discrimination, a lack of diversity will facilitate it.

33 Zoe Corbyn (2011), ‘Hungry judges dispense rough justice’, Nature, International Weekly Journal of Science. Available at: <https://www.nature.com/news/2011/110411/full/news.2011.227.html>



“There will be new jobs for humans to work out what machines are doing. And this is where it comes back to diversity – those humans in the loop must be diverse, so they can see the true range of possible impacts the machine is having.”

**Professor Dame Wendy Hall, Regius Professor of Computer Science, University of Southampton and co-author, UK government AI review**

Others suggested that public officials should also be expected to alter or limit the scope and powers of an AI system when it displays a high degree of bias. The impact of bias on the general public can be reduced, for example, by removing a system from the front-line of service delivery. Given the risk to both public trust and public standards, officials should be prepared to remove an AI system entirely if it persistently produces biased results.

While it was seen as implausible to prohibit any system displaying bias, public bodies should always know how their systems are biased and who is most affected by that bias. Once the bias of a system is known, suitable remedial action and mitigation procedures should follow.

“What we might want to say is ‘it is unacceptable not to know the ways in which your system is biased, and you are then required to account for how you use and understand the results of that system in that context.’ You need to be able to provide a justification and that justification has to be subject to scrutiny and challenge.”

**Oliver Buckley, Executive Director, CDEI**

Data scientists consulted for this review also outlined a number of technical methods that could be deployed to mitigate bias, while voicing scepticism that any AI system could be completely free of bias. As a matter of good practice, technologists recommended programming systems to exclude characteristics like race, gender, or age from predictive models. However, it was widely accepted that this would only have a limited impact. This was due to challenges around proxy characteristics. You could strip out ethnicity, for example, but location could act as an effective proxy, resulting in the same discriminatory effects.

“A draft tool we have looked at (at West Midlands Police) had intelligence information built in as input factors, including things like the number of stop and search counts, and that raised red flags around what that could be a proxy for in that particular region.”

**Marion Oswald, Senior Fellow in Law and Director of the Centre for Information Rights, University of Winchester**

Proxy characteristics can also be extremely subtle and not easily identifiable. A predictive policing model used to predict the likelihood of criminals reoffending could use natural language processing to analyse police interviews. That model could identify a defendant’s defensive response to questioning as an indicator of a propensity to reoffend. However, a defensive tone may instead be a response to a more aggressive line of questioning from the interviewing police officer, and police officers may, historically, have been more likely to ask more aggressive questions of male and ethnic minority interviewees.

Furthermore, stripping out certain characteristics may not make a system less biased. In the reoffending model outlined above, one answer could be to strip out gender as a likely predictor of reoffending rates. However, as the CDEI state, “Blinding algorithms to demographic differences and proxies for these differences does not always lead to fairer outcomes...Preventing an algorithm designed to calculate the risk of criminals reoffending from taking into account their sex, would likely result in disproportionately harsher sentences for women overall as women tend to reoffend less often than men. By excluding sex, the algorithm becomes less accurate for women and so, arguably, less fair.”<sup>34</sup>

“I’m not convinced that human cleansing of data adequately answers this problem. When we remove certain data points, how are we sure that we are making a dataset less biased? Whose rules are being used, why and who is saying that those rules are the right ones?”

**Sana Khareghani, Head, Office for AI**

Some suggested that a better solution was to increase the size and diversity of datasets. Overall, however, there was recognition that more research was needed into technical solutions to data bias.

---

34 Centre for Data Ethics and Innovation (CDEI) (2019), Interim report: Review into bias in algorithmic decision making, 11. Available at: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/819168/Interim\\_report\\_-\\_review\\_into\\_algorithmic\\_bias.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/819168/Interim_report_-_review_into_algorithmic_bias.pdf)



# Chapter 3: Guidance and ethical principles

## 3.1. Introduction

While AI poses particular risks and opportunities for openness, accountability and objectivity, there is nothing inherently new about what is needed to govern and manage AI responsibly. The Committee found that artificial intelligence does not necessitate a fundamental reworking of public sector practice. Successful AI governance is a question of clear regulation and proper controls for understanding, managing and mitigating risk.

In this sense, AI is a new challenge that can be solved with effective governance and a traditional risk management approach. The senior leadership of public bodies will first need to assess the risk an AI tool poses to public standards. They will then need to set governance mechanisms that mitigate that risk to a level deemed acceptable for the context AI is used in. Senior leadership will need to justify and be ultimately accountable for any risk mitigation measures their organisations take. By implementing the right processes, policies and management structures, public bodies will remain accountable, open and objective when using AI.

Public sector organisations will not, however, be able to establish sound governance alone. AI poses new challenges around issues such as explainability and responsibility that public sector organisations will not encounter when using conventional digital systems. Public bodies will need clear guidance based on sound ethical principles on how to adapt their governance and management structures for AI. To this end, the Office for AI, the Government Digital Service, and the Turing Institute collectively published *A Guide to Using Artificial Intelligence in the Public Sector* ('the AI Guide'), a comprehensive set of guidance for public bodies to use.<sup>35</sup>

Separately, the ICO has published its AI Auditing Framework. In collaboration with the World Economic Forum, the Office for AI has also produced specialist guidelines on AI procurement. This chapter assesses the quality, practicality and accessibility of guidance issued so far.

With the establishment of the Office for AI and the CDEI, and the designation of the Alan Turing Institute as the UK's national centre for AI, the UK government has signalled its ambition to become a world leader in responsible innovation. It should be noted that the issuing of the AI Guide is the most significant piece of work published towards this goal. However, the Committee's view is clear: guidance alone is not enough, and clear, well-established regulation is needed to ensure the responsible use of AI in the public sector. The form that AI regulation could take is discussed in chapter 4.

## 3.2. Ethical principles

"A hallmark of good governance is the development of shared values, which become part of the organisation's culture, underpinning policy and behaviour throughout the organisation, from the governing body to all staff."<sup>36</sup>

**The Independent Commission on Good Governance**

35 Government Digital Service, Office for AI and the Alan Turing Institute (2019), *A Guide to Using Artificial Intelligence in the Public Sector*. Available at: <https://www.gov.uk/government/collections/a-guide-to-using-artificial-intelligence-in-the-public-sector>

36 The Independent Commission on Good Governance (2004), *The Good Governance Standard for Public Services*, 13. Available at: <https://www.jrf.org.uk/report/good-governance-standard-public-services> as quoted by this Committee in its report *Standards Matter*, 22



Ethical principles underpin good public sector governance. AI is no exception. Establishing a clear set of ethical principles covering the use of AI in the public sector reminds public officials to consider public standards when using AI, and to choose a course of action that best adheres to those principles. As artificial intelligence will fundamentally change the way public services are delivered, the public sector needs a clear set of ethical principles specific to the challenges posed by AI.

The final section of the AI Guide, 'Using AI ethically and safely', begins on this basis, establishing a new set of values and principles, known as the FAST Track Principles and the SUM Values.<sup>37</sup> The SUM Values "support, underwrite, and motivate a responsible innovation ecosystem" by outlining the values that underpin the ethical permissibility of an AI project. Those values are respect, connect, care, protect. The FAST principles guide the design and use of AI systems. They are fairness, accountability, sustainability and transparency.

The establishment of ethical principles specifically for AI in the UK public sector is welcome. Academics estimate that over 70 AI ethics codes have been published over the past three years, and contributors emphasised the risk of 'ethics-shopping', where, as Professor Luciano Floridi argues, "private and public actors may shop for the kind of ethics that is best retrofitted to justify their current behaviours, rather than revising their behaviours to make them consistent with a socially accepted ethical framework."<sup>38</sup> A single statement of AI ethical values is a significant step forward in solving this problem.

The SUM Values provide a good starting point for public officials debating whether or not to introduce AI. AI will create new possibilities in prediction, automation and analysis, so it is important that public sector organisations examine the ethical permissibility of their project before deciding to procure or build an AI system. Multiple contributors warned against public bodies taking a "shiny new tool" approach to AI where projects were embarked on without consideration of their long-term social, ethical and environmental impact.

The FAST principles provide a clear, actionable and appropriate guide for public sector behaviour. Fairness, accountability, sustainability and transparency comprehensively cover the five areas of concern identified in chapter two of this review, and complement, rather than contradict, the Seven Principles of Public Life.

The FAST principles certainly need greater distribution and promotion across public life. Many contributors called for a 'Super-Code' of ethical principles and the FAST principles could provide this. However, in order for an overarching set of AI ethical principles to gain traction across the public sector, the principles should be promoted more prominently, and the descriptors should be shorter and clearer. The principles should also be made explicit in all sector-specific AI codes of conduct.

37 Dr David Leslie (2019), *Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector*. The Alan Turing Institute. Available at: [https://www.turing.ac.uk/sites/default/files/2019-06/understanding\\_artificial\\_intelligence\\_ethics\\_and\\_safety.pdf](https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf)

38 Luciano Floridi (2019), 'Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical', *Philosophy & Technology*, 32:185. Available at: <https://link.springer.com/article/10.1007/s13347-019-00354-x>



## FAST Track Principles



In order to become authoritative, the FAST principles must live outside of the AI Guide. In the way that the Seven Principles of Public Life are the defining mission of this Committee, upholding the FAST principles could become the overarching goal of the Office for AI, CDEI, and Turing Institute's public sector work. For principles to shape institutions, they need to be integrated into public sector cultures.

This is all the more urgent as the establishment of the FAST principles has not solved the problem of ethics shopping, outlined above. Currently there are three sets of ethical principles endorsed by UK government bodies. As well as the AI Guide, the Department for Digital, Culture, Media and Sport (DCMS) published the Data Ethics Framework (DEF), which prescribes a number of useful values and practices, while the Centre for Data Ethics and Innovation (CDEI) adheres to OECD Principles on Artificial Intelligence, which the government has also adopted. Each of these three sets has a different focus, and some are more high-level than others, but this multiplicity of principles and codes confuses the landscape and undermines attempts to make any set of ethical principles authoritative. It is also unclear how they work together. For example,

although the AI Guide mentions that its principles are intended to supplement the Data Ethics Framework, it is unclear how they work together in practice.

Elevating the reach and status of an authoritative set of principles is also necessary given the prominence of private companies in AI-enabled public service delivery. Private providers may have their own lists of ethical principles that are inappropriate for public service delivery, or may exploit ambiguities in the higher-level and less focused principles adopted by government bodies.

It is noted that the Data Ethics Framework is currently under review, and that the AI Guide is intended as an iterative document. The government should use this opportunity to identify, endorse and promote an authoritative high level set of ethical principles. The public should be able to find easily a clear statement of ethical principles that govern the use of AI in the public sector, and it should be made clear to those on the frontline of service delivery which ethical principles public officials are expected to adhere to. This should include outlining the purpose, scope of application and respective standing of each of the three sets currently in use.





### Recommendation 1a:

**There are currently three different sets of ethical principles intended to guide the use of AI in the public sector – the FAST SUM Principles, the OECD AI Principles, and the Data Ethics Framework. It is unclear how these work together and public bodies may be uncertain over which principles to follow.**

**The public needs to understand the high level ethical principles that govern the use of AI in the public sector. The government should identify, endorse and promote these principles and outline the purpose, scope of application and respective standing of each of the three sets currently in use.**

Clear and authoritative ethical principles then need to be further elaborated and specified in codes of conduct that are explicit about what is expected of public office-holders in different contexts. It is likely that sector-specific AI ethics codes will be necessary, particularly in high-risk policy areas such as policing, criminal justice, health and social care. Sector-specific codes can help make abstract ethical principles clearer and more tailored to particular professional settings, while retaining the link to the standards expected of public office-holders across the whole of the public sector.

The Code of Conduct for Data-Driven Health and Care Technology produced by the Department of Health and Social Care (DHSC) is a good example of best practice, which integrates AI ethics with pre-existing medical and public standards.<sup>39</sup>

Codes of conduct that elaborate what the principles imply in particular organisations ensure that everyone in the organisation knows what is expected of them. They also inform those holding them to account. This is useful where the application of principles may not be self-evident and where it remains unclear how public officials will uphold these ethical principles in practice. The DHSC code, for example, asks public office-holders to “Generate clear evidence of the effectiveness and economic impact of a product or innovation...an evidence-generation plan should be developed using the evidence standards framework published by the National Institute for Health and Care Excellence (NICE).”<sup>40</sup>

Such codes should not, however, override principles. Behaviour of public office-holders can technically be within the rules set out in a code and yet still offend against underlying principles and values expected of them by the public. Principles and codes must be viewed as complementary rather than alternatives.

Public officials should also be aware that ethics is an ongoing and dynamic practice. Principles and codes serve as a useful guide for those looking to make the right judgement in a particular context, but they are not a substitute for comprehensive ethical risk management. Government departments should be aware that establishing and promulgating an AI ethics code is the beginning, and not the end, of effective AI governance.

39 Department of Health and Social Care (2019), Code of conduct for data-driven health and care technology. Available at: <https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology>

40 Same source



### 3.3. The Office for AI, GDS and the Turing Institute’s ‘Understanding artificial intelligence ethics and safety’ – The AI Guide

Best practice and guidance is a vital part of any framework designed to uphold ethics. Public bodies will need guidance that shows how ethical principles for AI translate into practice. The Office for AI, the Government Digital Service (GDS), and the Turing Institute published the AI Guide in August 2019. The guidance helps public bodies understand AI, then covers three stages of an AI project, and ends with a longer document on AI ethics.

This guidance is welcome, and it is a comprehensive and valuable resource for public bodies wanting to implement AI while upholding high public standards.

#### **A Guide to Using Artificial Intelligence in the Public Sector**

- Understanding artificial intelligence
- Assessing if AI is the right solution for your users’ needs
- Planning and preparing for artificial intelligence implementation
- Managing your artificial intelligence project
- Using artificial intelligence ethically and safely.<sup>41</sup>

Importantly, ethical standards are not restricted to the final section of the AI Guide. The guidance emphasises that ethics must be considered at every stage of an AI process, from assessing if AI is the right solution, through project planning, to system management. Public bodies using this guidance should ensure they follow every section, rather than the section on ethics and safety alone.

The section most relevant to this review is ‘Using artificial intelligence ethically and safely’, produced by the Turing Institute and summarised on the Office for AI website. This guidance reflects a number of key points made by contributors to this review: AI ethics (and therefore ethics-based governance) is heavily context-specific; ethical principles must be actionable; and ethics-based governance is a continuous process, rather than a one-time event. The guide’s integration of ethical issues into a process-based governance (PBG) framework is laudable, and reflects a core conclusion of this review: that high public standards are a product of good governance.

It remains to be seen if the guidance will have a significant impact on AI in the public sector. We were informed that future iterations of the AI Guide would be subject to more extensive publicity. This is vital. Guidance, no matter how good, will leave no mark on the landscape without extensive measures to promote its adoption.

Future iterations of ‘Using artificial intelligence ethically and safely’ must also be made easier to use and understand. The full document is nearly 100 pages in length and assumes a level of technical awareness above what can be reasonably expected of senior leadership in a local council, school or police force. This undermines the practicality of the guidance, especially as leadership – those setting governance – is the intended audience for this guide. In its current form, implementing the guidance would require oversight from specialist AI policy professionals, but this is not a resource many public sector organisations will have. As it stands, the ethics and safety guidance would work better as a source document for sector-specific guidance and best practice, rather than an authoritative guide for all public sector organisations to follow.

41 Government Digital Service, Office for AI and the Alan Turing Institute (2019), A Guide to Using Artificial Intelligence in the Public Sector. Available at: <https://www.gov.uk/government/collections/a-guide-to-using-artificial-intelligence-in-the-public-sector>



Good guidance has the potential to change behaviours and shape professional cultures. It is always an important part of any standards regime. Guidance is, however, a non-binding tool. The issuing of good guidance does not constitute the formalisation of public standards for AI. Using Artificial Intelligence in the Public Sector is a high-quality set of guidance, as is ‘Using artificial intelligence ethically and safely’. But further work on promotion and accessibility is needed to ensure this guidance has the greatest effect.

#### **Recommendation 1b:**

**The guidance by the Office for AI, the Government Digital Service and the Alan Turing Institute on using AI in the public sector should be made easier to use and understand, and promoted extensively.**

Contributors did question if the Office for AI and GDS were the right organisations to issue guidance, given a potential conflict of interest between promoting AI adoption and upholding ethical standards.

“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 organisation has the capacity or remit to ensure robust and consistent ethical supervision on broader questions of automated decision system adoption and use in public policy, including their use outside central government.”

**Dr Emma Carmel, Associate Professor, Social and Policy Sciences, University of Bath**

This concern, while valid, does not fully reflect the nature of the Office of AI and GDS, both of which have shown a clear commitment to ethics as well as adoption. This is reflected in the guidance, which makes clear that AI is a data science tool of limited utility for addressing specific problems, and not a universal solution to any public policy challenge.

#### **3.4. The ICO’s auditing framework for AI**

The ICO has also issued guidance relevant to public standards and AI through its auditing framework, which provides information on how data processors can ensure compliance with data protection requirements for AI under the GDPR.<sup>42</sup> It covers issues of fairness and transparency that mirror this Committee’s concerns.

42 ICO (2019), An overview of the Auditing Framework for Artificial Intelligence and its core components. Available at: <https://ico.org.uk/about-the-ico/news-and-events/ai-blog-an-overview-of-the-auditing-framework-for-artificial-intelligence-and-its-core-components/>



Explanations and guidance on each individual risk area is given in a series of blog posts. These are useful, provide clear instructions on how to mitigate risk, and are written in an accessible way for those without technical expertise. It is an important resource for public bodies wanting to uphold public standards in the areas covered. ICO blog posts on fully automated systems and bias and discrimination provide the best user-orientated guidance on these topics seen in the course of this review.

In its analysis of fully automated decision-making models, ICO guidance states:

- human reviewers must be involved in checking the system’s recommendation and should not “routinely” apply the automated recommendation to an individual
- reviewers’ involvement must be active and not just a token gesture. They should have actual “meaningful” influence on the decision, including the “authority and competence” to go against the recommendation
- reviewers must “weigh-up” and “interpret” the recommendation, consider all available input data, and also take into account other additional factors.<sup>43</sup>

### Proposed framework

1. Governance and accountability				
Cross-cutting focus areas	Risk appetite	Leadership engagement and oversight	Management and reporting structures	Compliance and assurance capabilities
	Data protection by design and by default	Policies and procedures	Documentation and audit trails	Training and awareness
2. AI-specific risk areas				
Fairness and transparency in profiling	Accuracy	Fully automated decision making models	Security and cyber	
Trade-offs	Data minimisation and purpose limitation	Exercise of rights	Impact on broader public rights*	

\*Includes only considerations with scope of an ICO investigation/audit

43 ICO (2019), Automated Decision Making: the role of meaningful human reviews. Available at: <https://ico.org.uk/about-the-ico/news-and-events/ai-blog-automated-decision-making-the-role-of-meaningful-human-reviews/>

Public sector organisations should be aware that the auditing framework for AI is constrained by the ICO's remit, and that the GDPR is not a perfect fit for all public standards concerns. In particular, contributors cautioned against public sector organisations assuming that a Data Protection Impact Assessment covers all ethical issues; specialist AI impact assessment will also be necessary (see chapter 4). Nonetheless, public sector organisations should be confident that the auditing framework provides an authoritative steer on the ethical issues covered by the GDPR, including explainability, automation and responsibility, and bias.<sup>44</sup>

### 3.5. The Office for AI's guidelines for AI procurement

The Draft Guidelines for AI Procurement published by the Office for AI and the World Economic Forum in September 2019 outline a number of useful ways for public bodies to compel private providers of public services to consider ethics. These include referencing ethical principles such as the government's Data Ethics Framework in invitations to tender, developing strategies to address the ethical limitations of training data, and using ethical considerations as evaluation criteria.<sup>45</sup> The guidelines are a work in progress and undergoing trial at the time of writing.

There are key robust practices you can ask for suppliers to demonstrate when providing AI solutions. The guidance for understanding AI ethics and safety provides a useful framework to identify those. Besides having an ethical framework within their company, robust practices include:

- having an internal AI ethics approach, with examples of how it has been applied to design, develop, and deploy AI solutions
- processes to ensure accountability over outputs of algorithms
- avoiding outputs of analysis which could result in unfair decision making

The Office for AI's Draft Guidelines for AI procurement.<sup>46</sup>

44 ICO (2019), Project Explain, Interim report. Available at: <https://ico.org.uk/media/about-the-ico/documents/2615039/project-explain-20190603.pdf>

45 Office for AI (2019), Draft Guidelines for AI Procurement. Available at: <https://www.gov.uk/government/publications/draft-guidelines-for-ai-procurement/draft-guidelines-for-ai-procurement>

46 Same source



The Office for AI's guidelines also helpfully outline how public bodies can reform their own internal practices to ensure that the people commissioning the technology understand the importance of considering public standards, such as transparency and accountability, throughout the procurement process. Their guidance on building multidisciplinary teams is particularly useful. The Committee frequently heard how important it is to combine the expertise of data scientists, data ethics specialists and policy experts when using AI technology, and it is likely that "developing, evaluating and delivering AI invitation-to-tenders will be more effective with diverse teams that understand the interdependent disciplines AI covers."<sup>47</sup> The guidelines' extensive use of the Data Ethics Framework (DEF) is also welcome, as this ensures consistency and continuity across an AI project for public sector organisations using the DEF in AI deployment.

Private providers of public services are subject to the Seven Principles of Public Life, and this area has been the focus of the Committee's attention before. As this past work has shown, when awarding contracts public bodies should consider the ethical behaviour and culture of a company, as well as whether the AI product meets ethical standards. The procurement process should be used to convey to private companies that they have ethical obligations throughout the entire course of a contract and that ethics is not a one-off event, nor one that can be devolved to the public sector purchaser. The Committee's recommendations for reform of the procurement process are discussed in chapter 4.

---

47 Same source



# Chapter 4: Regulating AI

## 4.1. Introduction

Recent guidance published by the Office for AI (in partnership with the Turing Institute and GDS) and the ICO marks a welcome and significant step forward in AI governance across the public sector. This guidance, covered in chapter 3, provides a good starting point for thinking about how public bodies can establish process-based governance mechanisms that safeguard public standards when they are using AI. The Committee believes, however, that guidance alone does not provide a strong enough incentive to change behaviour.

A strong and coherent regulatory framework for AI in the UK public sector is still a work in progress. A comparison between AI in healthcare and AI in policing is instructive. Healthcare practitioners told the Committee they were confident AI could be implemented ethically because medicine operates within a strictly regulated system, where there is already in place a professional system for testing, integrating and challenging new practices and technologies, and clear standards for reporting, research and clinical trials. Experts working in the field of medical AI told this review that new technologies would slot easily into this pre-existing framework.

In contrast, the same established and well-understood regulatory framework does not currently exist in policing. There is no clear process for evaluating, procuring or deploying new technologies such as predictive policing or facial recognition, which are already being used to support decision-making across the UK. In the absence of a clear regulatory framework for policing, safeguards for public standards are left to individual police forces, whose recent attempts at creating ethical

AI systems have led to mixed results.<sup>48</sup> Evidence submitted to this review showed that the use of AI in policing is far more representative of the wider public sector than AI in healthcare. AI may be used in areas such as education, social care and welfare, without a proper understanding of the distinctive value added or risks created by AI systems, their impact on citizens, and the extent to which they serve legitimate policy aims.<sup>49</sup> Hence the need for a strong regulatory framework.

Efforts to establish clear regulation for AI are underway. The General Data Protection Regulation 2018 (GDPR) establishes an extensive legal framework for any organisation processing personal data, including provisions for automated processing. Through its strong ethical foundation and fair processing requirements, the GDPR safeguards against many of the standards issues highlighted in this report. The ICO, as the UK's data protection regulator, is currently looking at how the GDPR applies to AI. Their conclusions will form a substantive part of the UK's regulatory landscape for AI.

Other laws, regulations and public bodies are also relevant. The Equality Act 2010 prohibits discrimination against certain protected characteristics, making it the key law safeguarding against data bias. The Centre for Data Ethics and Innovation (CDEI) was established to advise government on AI regulation. Procurement processes act as a form of soft regulation, setting the terms for commercial relationships. Mandatory impact assessments can change public sector behaviour and obligations under the Freedom of Information Act set the terms for transparent disclosure.

48 In October 2019, the ICO issued a formal Opinion on Live Facial Recognition. The Information Commissioner found the current laws, codes and practices relating to LFR will not drive the ethical and legal approach needed to manage the risk.

ICO (2019), Information Commissioner's Opinion: the use of live facial recognition technology by law enforcement in public places. Available at: <https://ico.org.uk/media/about-the-ico/documents/2616184/live-frt-law-enforcement-opinion-20191031.pdf?hootPostID=d672132320a2e1a6fea681db20056c9>

49 Alexander Babuta, Marion Oswald and Christine Rinik (2018), 'Machine Learning Algorithms and Police Decision-Making: Legal, Ethical and Regulatory Challenges', RUSI Whitehall Report, 3-18. Available at: [https://rusi.org/sites/default/files/201809\\_whr\\_3-18\\_machine\\_learning\\_algorithms.pdf.pdf](https://rusi.org/sites/default/files/201809_whr_3-18_machine_learning_algorithms.pdf.pdf)



The Committee has concluded, however, that even taken together this regulatory framework is not yet fit for purpose. Though improvements have been made in recent months and years, current regulation – as it is understood and implemented across the public sector – does not provide a strong enough defence against the risks to public standards identified in this report. This chapter hopes to provide some direction on how government should regulate AI to uphold public standards, covering in turn the GDPR; the Equality Act; the CDEI; Procurement and the Digital Marketplace; Impact Assessment; and Transparent Disclosure. This chapter is not a comprehensive examination of AI regulation. It is limited to the areas that most directly affect the three public standards at the core of this review: openness, accountability and objectivity.

## 4.2. Legal compliance

Any effective system of public sector regulation requires public bodies to take proactive measures to comply with existing legislation and ensure there is a clear basis in law for any activity they undertake. However, there was a widespread perception among contributors to this review that public bodies are introducing AI into service delivery without a clear understanding of the requirements of the law. Concerns were most pressing in law enforcement and the judiciary, where new surveillance capabilities, such as automated facial recognition (AFR), will impact on citizens' rights and freedoms.

Legal experts told the Committee that public bodies were often relying on a tenuous and piecemeal legal basis, often constituted from multiple sources, to legitimate the use of new technology. Contributors criticised the fact that intrusive and controversial technology, which has the potential to reshape society in radical ways, is introduced in this way.

“[It is] 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.”<sup>50</sup>

**Professor Karen Yeung, Interdisciplinary Professorial Fellow in Law, Ethics and Informatics, University of Birmingham Law School and School of Computer Science**

The validity of these legal bases is already being tested in the courts. In September 2019, the High Court found that the use of facial recognition by South Wales Police was lawful.<sup>51</sup> Some contributors welcomed the use of judicial review to establish legal clarity, viewing it as an important mechanism to establish checks and balances on executive power. Others, however, argued that it is not appropriate for the legislative framework for era-defining technology to be created by judicial review, especially when much of the legislation subject to review was not designed with AI in mind.

Public bodies should not implement AI without understanding the legal framework governing its use. Introducing algorithmic systems into the public sector without a clear legal basis not only undermines public standards, but also the rule of law. Judicial review may create legal clarity but a series of high-profile court cases investigating illegality by public bodies will undermine trust in what can be a potentially beneficial technology.

50 Written evidence 20 (Professor Karen Yeung)

51 *Bridges v Chief Constable of South Wales Police* [2019] EWHC 2341. Available at: <https://www.judiciary.uk/wp-content/uploads/2019/09/bridges-swp-judgment-Final03-09-19-1.pdf>





The Law Society, in its report on the use of algorithms in the criminal justice system, recommended that “[t]he lawful basis of all algorithmic systems in the criminal justice system must be clear and explicitly declared in advance.”<sup>52</sup> This should apply not only to the criminal justice system, but to the public sector in general. Public bodies should publish a statement on how their use of AI complies with the relevant laws and regulations before they are deployed in public service delivery.

#### **Recommendation 2:**

**All public sector organisations should publish a statement on how their use of AI complies with the relevant laws and regulations before they are deployed in public service delivery.**

### **4.3. The GDPR**

Given that most uses of AI in the public sector will involve the processing of citizens’ personal data, the GDPR – which has direct application in UK law through the Data Protection Act 2018 – creates an extensive legal framework for AI. It places a number of obligations on organisations handling personal data and has a strong ethical foundation. The GDPR gives people enhanced protections against unnecessary data collection, and seeks to limit the intrusive use of data through its principles of fairness and privacy by design, which in turn protect a range of further rights.

Article 5 of the GDPR sets out key principles for the processing of personal data. These include lawfulness, fairness and transparency; purpose limitation; accuracy; and accountability. Many of these normal obligations are risk-based and especially pertinent to AI. Insofar as automated

decision-making involves the processing of personal data, all of these provisions apply. For example, organisations must identify a lawful basis for collecting and processing personal data. An organisation that does not establish a clear legal basis for the use of AI would not only undermine public standards but would likely be in breach of data protection legislation.

Full knowledge and articulation of purposes for processing are also required by the purpose specification and use limitation principles. These say that personal data should only be collected for specified purposes and then only used for those purposes or purposes that are compatible with the original one. This could provide an effective safeguard for ensuring that AI is only used for the purpose it is meant to serve. However, it is likely that a narrow interpretation of this principle may not prove useful, particularly because AI may yield unforeseen and sometimes unpredictable results.

Data protection law is technology neutral. It does not directly refer to AI or any associated technologies such as machine learning. However, the GDPR does have a significant focus on large scale automated processing of personal data, and several provisions make specific reference to the use of profiling and automated decision-making. This means that it applies to the use of AI to provide a prediction or recommendation about someone. For example, the law requires organisations to handle personal data in ways that people would reasonably expect and not in ways that are unduly detrimental and might cause harm. This would likely require organisations to use AI in ways that are proportionate and not discriminatory.

52 The Law Society (2019), Algorithms in the Criminal Justice System, 61 (Recommendation 5). Available at: <https://www.lawsociety.org.uk/support-services/research-trends/algorithm-use-in-the-criminal-justice-system-report/>



---

<b>The right to be informed</b>	Articles 13 and 14 of the GDPR give individuals the right to be informed of the existence of solely automated decision-making, meaningful information about the logic involved, and the significance and envisaged consequences for the individual.
<b>The right of access</b>	<p>Article 15 of the GDPR gives individuals the right of access to information on the existence of solely automated decision-making, meaningful information about the logic involved, and the significance and envisaged consequences for the individual.</p> <p>Recital 71 provides interpretive guidance. It says that individuals should have the right to obtain an explanation of a solely automated decision after it has been made, but it is not legally binding.</p>
<b>The right to object</b>	Article 21 of the GDPR gives individuals the right to object to processing of their personal data, specifically including profiling, in certain circumstances.
<b>Rights related to automated decision-making including profiling</b>	<p>Article 22 of the GDPR gives individuals the right not to be subject to a solely automated decision producing legal or similarly significant effects. There are some exceptions to this and in those cases it obliges organisations to adopt suitable measures to safeguard individuals, including the right to obtain human intervention, to express their view, and to contest the decision.</p> <p>Recital 71 also provides guidance for Article 22.<sup>53</sup></p>

---

The GDPR is enforced by a specialist data protection regulator, the Information Commissioner's Office (ICO). The ICO also has a number of enforcement powers, which help to safeguard against potential breaches of data protection legislation. The ICO can impose penalty notices, fine organisations up to £20 million or 4% of their annual turnover (whichever is higher) for breaking the law, and issue guidance that must be considered by Courts arbitrating on the GDPR.

The ICO also issues formal opinions. These opinions, though non-binding, carry significant weight and authority and should encourage organisations to comply with the views of the Commissioner. For example, in October 2019 the Information Commissioner issued an opinion on the use of live facial recognition technology (LFR) by law enforcement in public places,<sup>54</sup> following the High Court judgement on South Wales Police

Force. It found that sensitive processing happens at each stage of the LFR process and as such it is subject to data protection law, including the EU Law Enforcement Directive.

Issues of openness, responsibility, explanations and accountability examined in this review are all covered by the GDPR. Overall, the GDPR regulates these issues well. Though some legal experts voiced doubts that the law covered the issues of responsibility and explainability sufficiently, the Committee is satisfied that ICO guidance resolves the issues identified.

#### 4.3.1. The GDPR and openness

Articles 13, 14 and 15 of the GDPR cover several elements of openness. Articles 13 and 14 are transparency obligations. These Articles tell organisations what information they must disclose to individuals before processing their data.

---

53 ICO and the Alan Turing Institute (2019), Explaining decisions made with AI, Draft guidance for consultation. Available at <https://ico.org.uk/media/about-the-ico/consultations/2616434/explaining-ai-decisions-part-1.pdf>

54 ICO (2019), Information Commissioner's Opinion: the use of live facial recognition technology by law enforcement in public places. Available at: <https://ico.org.uk/media/about-the-ico/documents/2616184/live-frt-law-enforcement-opinion-20191031.pdf?hootPostID=d67213232a0a2e1a6fea681db20056c9>



This includes the purpose of and lawful basis for processing, and also details of the existence of automated decision-making, including profiling.

These provisions effectively say that if you use AI to make solely automated decisions about people with legal or similarly significant effects, you must tell them what information you use, why it is relevant and what the likely impact is going to be.<sup>55</sup> Article 15 has similar effects but has to be triggered by the data subject. It says that individuals have a right to access information about the processing of their personal data after such processing has taken place, including where solely automated decision-making systems were used.<sup>56</sup> These rights, taken together, provide a clear regulatory obligation for public sector organisations to be transparent about their use of AI.

#### 4.3.2. The GDPR and responsibility

At first glance, Article 22 imposes a general restriction on “solely automated decision-making” and profiling where it results in a decision with “legal or similarly significant effects”. In theory this would prevent public bodies from implementing AI where no human has intervened in the decision-making process, creating a strong legal safeguard against the removal of human responsibility in a public sector decision-making process.

Some legal experts told the Committee, however, that Article 22 is less of a safeguard than it appears to be. This is because the word “solely” effectively undermines the provision, as it would permit any automated system subject to a cursory glance by a human operator, even if the human operator did not or could not make any changes or contribute to the operation of the system. The law could allow public officials to circumvent these provisions by rubber-stamping AI decisions with little or no human intervention in the decision-making process.

Data protection experts told the Committee that the applicability of the provision could be improved by using the phrase “solely or predominantly based on”, or by using a more detailed definition of automated decision-making, where the nature and type of human involvement is specified.<sup>57</sup>

However, ICO guidance makes clear that a public official automatically approving an AI decision does not constitute sufficient human involvement in the decision-making process. This interpretation is supported by the EU’s Article 29 Data Protection Working Party guidelines on automated decision-making, which say that the data controller cannot avoid Article 22 provisions by fabricating human involvement. For example, if someone inputs data to be processed, but has no influence on the decision, it may still be considered solely automated.<sup>58</sup> Given that ICO guidance must be considered by the courts in any AI cases, the Committee is of the view that the law as it currently stands provides an adequate safeguard against fully automated decision-making in the public sector.

55 ICO [online], Guide to the General Data Protection Regulation: The right to be informed. Available at: <https://ico.org.uk/media/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/the-right-to-be-informed-1-0.pdf>

56 ICO [online], Guide to the General Data Protection Regulation: The right of access. Available at: <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-of-access/>

57 Sandra Wachter, Brent Mittelstadt, Luciano Floridi (2017), ‘Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection’, *International Data Privacy Law*, 7;2. Available at: <https://academic.oup.com/idpl/article/7/2/76/3860948>

58 Article 29 Data Protection Working Party (2018), Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, IV, 21. Available at: [https://iapp.org/media/pdf/resource\\_center/W29-auto-decision\\_profiling\\_02-2018.pdf](https://iapp.org/media/pdf/resource_center/W29-auto-decision_profiling_02-2018.pdf)



“[H]uman involvement has to be active and not just a token gesture. The question is whether a human reviews the decision before it is applied and has discretion to alter it or whether they are simply applying the decision taken by the automated system.”<sup>59</sup>

### **What does the GDPR say about automated decision-making and profiling? ICO**

#### **4.3.3. The GDPR and explanations**

There is legal uncertainty around the right to explanation, which is said to exist under the GDPR. If this provision were to exist, it would grant citizens a legally mandated and meaningful right to explanation for decisions made by automated systems.

This would be a promising legal mechanism in the broader pursuit by government of accountability and transparency in AI-enabled public service delivery. However, some legal experts told the Committee that such a right is unlikely to exist because there is nothing in the legally binding provisions of the GDPR that mandates a right to an explanation, and the idea of a right to an explanation only exists in non-binding recitals to the law. This means the law, in this regard, runs the risk of being toothless.

To provide guidance on explanations and clarify the law, the ICO and the Turing Institute are undertaking Project ExplA/n. In their guidance, they take an alternative view, stating that “the reference to an explanation of an automated decision after it has been made in Recital 71 makes clear that such a right is implicit in Articles 15 and 22.” Contributors to this review also emphasised that administrative law and the right to an appeal in UK law creates a strong legal incentive to provide an explanation for any public sector decision.

“You need to be able to give an individual an explanation of a fully automated decision to enable their rights, to obtain meaningful information, express their point of view and contest the decision.”<sup>60</sup>

### **ICO Guidance, Why Explain AI, Project ExplA/n**

The Committee is satisfied that ICO guidance provides a sufficient regulatory safeguard for the provision of explanations in public sector decision-making. ICO guidance should be considered authoritative and public bodies should provide explanations accordingly.

#### **4.3.4. The GDPR and accountability**

The GDPR includes its own explicit accountability principle, which says that organisations are responsible for the way that they use personal data and must have in place appropriate mechanisms for demonstrating compliance with GDPR principles. Article 24 of the GDPR says that organisations need to implement technical and organisational measures that are risk-based and proportionate to meet the requirements of accountability. Organisations are advised and in some cases required to: implement data protection policies; take a “data protection by design” approach; document processing activities; and carry out data protection impact assessments (DPIAs). AI that processes personal data will have to comply with these requirements, and the GDPR therefore provides a strong regulatory impetus for organisational accountability.

59 ICO [online], Guide to the General Data Protection Regulation: What does the GDPR say about automated decision-making and profiling? Available at: <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/automated-decision-making-and-profiling/what-does-the-gdpr-say-about-automated-decision-making-and-profiling/>

60 ICO and the Alan Turing Institute (2019), Explaining decisions made with AI, Draft guidance for consultation. Available at <https://ico.org.uk/media/about-the-ico/consultations/2616434/explaining-ai-decisions-part-1.pdf>



Data impact assessments are used to analyse, identify and minimise data protection risks. Irrespective of whether there is a new formal mechanism for AI risk assessment (see section 4.7), a DPIA is almost always going to be mandatory where public bodies are using AI to make decisions. This is because Article 35(1) says that organisations must carry out a DPIA where the type of processing is likely to result in a high risk to the rights and freedoms of individuals. This includes profiling, the large scale use of sensitive data and public monitoring, and is likely to include most, if not all, processing of personal data by innovative technology.<sup>61</sup>

#### 4.4. The Equality Act

Data bias could cause AI to produce decisions and policy outcomes that are discriminatory. Civil rights groups have criticised predictive policing models in particular, fearing that the use of AI could introduce discriminatory practice. Decisions may be made by algorithm without due consideration to policies and practices intended to safeguard those with protected characteristics, enhance diversity and improve outcomes for marginalised people. From a standards perspective, there is no reason to view discrimination resulting from biased data differently from discrimination resulting from human bias. Both undermine the Nolan Principle of objectivity.

“Although predictive policing is simply reproducing and magnifying the same patterns of discrimination that policing has historically reflected, filtering this decision-making process through complex software that few people understand lends unwarranted legitimacy to biased policing strategies that disproportionately focus on BAME and lower income communities.”<sup>62</sup>

#### **Policing by Machine, Liberty**

Biased decision-making may also violate non-discrimination law. The Equality and Human Rights Commission (EHRC) has statutory powers to enforce the Equality Act 2010, which prohibits discrimination against nine protected characteristics.<sup>63</sup> The Act also established the Public Sector Equality Duty (PSED) in 2011, which mandates public bodies to take a proactive approach to fighting inequality. There is nothing in the Equality Act that specifically refers to AI or automated decision-making. However, evidence from anti-discrimination lawyers outlined a number of ways in which the law’s provisions against direct and indirect discrimination could apply to AI.

61 ICO [online], Guide to the General Data Protection Regulation: When do we need to do a DPIA? Available at: <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/data-protection-impact-assessments-dpias/when-do-we-need-to-do-a-dpia/>

62 Liberty (2019), Policing by Machine: predictive policing and the threat to our rights. Available at: <https://www.libertyhumanrights.org.uk/sites/default/files/LIB%2011%20Predictive%20Policing%20Report%20WEB.pdf>

63 Equality Act 2010, section 4. Available at: <http://www.legislation.gov.uk/ukpga/2010/15/section/4>  
Protected characteristics defined by the Act are: age; disability; gender reassignment; marriage and civil partnership; pregnancy and maternity; race; religion or belief; sex; and sexual orientation



“In 2017, Durham Constabulary started to implement a Harm Assessment Risk Tool (HART), which utilised a complex machine learning algorithm to classify individuals according to their risk of committing violent or non-violent crimes in the future. This classification is created by examining an individual's age, gender and postcode. This information is then used by the custody officer, so a human decision maker, to determine whether further action should be taken. In particular, whether an individual should access the Constabulary's Checkpoint programme which is an “out of court” disposal programme.

There is potential for numerous claims here. A direct age discrimination could be brought by individuals within certain age groups who were scored negatively. Similarly, direct sex discrimination claims could be brought by men, in so far as their gender leads to a lower score than comparable women. Finally, indirect race discrimination or direct race discrimination claims could be pursued on the basis that an individual's postcode can be a proxy for certain racial groups. Only an indirect race discrimination claim would be susceptible to a justification defence in these circumstances.”<sup>64</sup>

#### AI Law Hub

Contributors to the review mentioned that the Public Sector Equality Duty (PSED), if used properly, was the single best tool available to deal with data bias. This is because it requires organisations to consider how they could positively contribute to the advancement of equality, and requires “equality

considerations to be reflected in the design of policies and delivery of services”.<sup>65</sup> Equality Impact Assessments are not required by law, but are often used by public bodies to facilitate compliance with the PSED. They are used to identify possible negative impacts of decisions on individuals and groups with protected characteristics and plan mitigating action accordingly. They are also used to identify opportunities to advance equality within the policies, strategies and services of a public authority.

“Public bodies must consider the Public Sector Equality Duty when they make decisions about how they fulfil their public functions and deliver their services. When moving towards automated decision making the PSED provides an opportunity for equality considerations to be built into decision-making processes as they are developed.”

**Rebecca Hilsenrath, Chief Executive,  
Equality and Human Rights Commission**

However, there is uncertainty around how the legislation applies in practice to automated decision-making in the public sector. There is currently no bespoke regulatory guidance outlining what public bodies introducing AI systems need to do to comply with the Equality Act 2010. Public bodies introducing AI systems need to know how the Act applies to discriminatory outcomes enabled by automated decision-making. They need specific guidance on how to comply with the legislation, as well as guidance on how to measure bias and mitigate its effects, particularly given the widespread belief among AI experts that data bias cannot or should not be completely eradicated.

64 Robin Allen QC and Dee Masters [online], ‘UK's existing equality and human rights framework’, AI Law Hub.

Available at: <https://ai-lawhub.com/framing-the-debate/#criminal>

On Durham Constabulary's HART model see: Alexander Babuta (2018), ‘Innocent Until Predicted Guilty? Artificial Intelligence and Police Decision-Making’, RUSI Newsbrief, 38;2.

Available at: [https://rusi.org/sites/default/files/20180329\\_rusi\\_newsbrief\\_vol.38\\_no.2\\_babuta\\_web.pdf](https://rusi.org/sites/default/files/20180329_rusi_newsbrief_vol.38_no.2_babuta_web.pdf)

65 Equality and Human Rights Commission [online], Public Sector Equality Duty Guidance. Available at: <https://www.equalityhumanrights.com/en/advice-and-guidance/public-sector-equality-duty>



Through Project ExplAI, the ICO and the Turing Institute are developing extensive guidance on explanations and the GDPR. The Committee believes that a similar project is necessary on data bias and the Equality Act. The Equality and Human Rights Commission should develop guidance on data bias in partnership with the Turing Institute and the CDEI.

### **Recommendation 3:**

**The Equality and Human Rights Commission should develop guidance in partnership with both the Alan Turing Institute and the CDEI on how public bodies should best comply with the Equality Act 2010.**

Though this project should focus on the Equality Act as it currently stands, some contributors suggested that a fundamental rethink of anti-discrimination law may be needed in the long term. The use of machine learning raises new issues that current anti-discrimination law may not cover. How will we detect cases of discrimination when a citizen may not even know if a decision has been made on the basis of a protected characteristic? Can discrimination law have any effect if discrimination occurs via proxy characteristics but we cannot identify what those proxies are? What forms of algorithmic profiling count as discrimination? Government should remain open to a revision of anti-discrimination law if the current legal framework cannot answer these questions convincingly.

### **4.5. Regulatory assurance body**

Some contributors to this review suggested that a new system of ethical regulation for the use of AI in the public sector was necessary. The Committee heard that a statutory arms-length public body, similar to the Human Fertilisation and Embryology Authority (HFEA), could have a role in licensing technology and leading on standards, review and assessment.<sup>66</sup> Justice of The Supreme Court, The Right Hon Lord Sales also called for an independent regulator of algorithms that would be staffed by technical experts, lawyers and ethicists. He argued that issues around AI are so large and impenetrable that an expert commission on algorithms is necessary to safeguard against the legal and ethical challenges posed by AI. Lord Sales said that this is particularly pertinent because government currently lacks the technical capacity to do this well itself.<sup>67</sup>

The Committee agrees with the rationale for extra regulatory scrutiny and independent advice on the issues associated with AI. However, most contributors to this review argued that a single AI regulator was impractical. The Committee heard that any system of ethical regulation for AI in the public sector would require sectoral-based review to account for the context specific risks and opportunities of automated decision-making across policy areas. A new AI regulator would inevitably overlap with existing regulatory bodies, who will already have to regulate AI within their sectors and remits. As such, the Committee believes that the UK does not need a new regulator.

66 Written evidence 12 (Dr Emma Carmel)

67 Lord Sales, Justice of the UK Supreme Court (2019), 'Algorithms, Artificial Intelligence and the Law', The Sir Henry Brooke Lecture for BAILII. Available at: <https://www.supremecourt.uk/docs/speech-191112.pdf>



“People often say ‘Let’s have a new regulator. Let’s have a new, shiny one.’ Actually, there is a lot of expertise already in the regulators because they are having to deal with this kind of thing in markets which they are there to regulate. We ought to build on that and use the expertise we have got.”

**Professor Helen Margetts, Professor of Society and the Internet, University of Oxford and Director of the Public Policy Programme, The Alan Turing Institute**

Instead, the Committee is of the view that existing regulators should be aware of how automated technology will impact their sectors, and adapt their practices accordingly. However, given the complexity of this technology and that expertise is not necessarily well established in this area, it is unlikely that regulators will be able to meet the challenges posed by AI without guidance from a central body. AI will create unforeseen issues for regulation, where technical knowledge and expertise will be necessary. There is clear space in the regulatory landscape for a “regulatory assurance” body, which provides advice to individual regulators and government on the issues associated with AI, and identifies any regulatory gaps. This body would not act as a regulator, but it would need full independence from government to advise objectively and without political interference.

The Centre for Data Ethics and Innovation (CDEI) has many of the necessary skills to fulfil this role and the Committee supports the government’s published intention for CDEI to oversee the regulatory landscape, analysing and anticipating gaps in governance and regulation that could impede the ethical deployment of AI, and to advise government accordingly.<sup>68</sup> The Committee also believes that the CDEI has a role to play in advising individual regulators, as well as government, on the issues associated with this technology. The Committee supports the government’s intention to place the centre on a statutory footing to safeguard its independence.<sup>69</sup> However, the specific roles and functions of the CDEI remain unclear. The government must clarify its purpose and assure that appropriate safeguards are in place so that it can fulfil its intended role as a regulatory assurance body.

68 Centre for Data Ethics and Innovation, Government Response to Consultation, November 2018. Available at: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/757509/Centre\\_for\\_Data\\_Ethics\\_and\\_Innovation\\_-\\_Government\\_Response\\_to\\_Consultation.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/757509/Centre_for_Data_Ethics_and_Innovation_-_Government_Response_to_Consultation.pdf)

69 Same source





**Recommendation 4:**

**Given the speed of development and implementation of AI, we recommend that there is a regulatory assurance body, which identifies gaps in the regulatory landscape and provides advice to individual regulators and government on the issues associated with AI.**

**We do not recommend the creation of a specific AI regulator, and recommend that all existing regulators should consider and respond to the regulatory requirements and impact of the growing use of AI in the fields for which they have responsibility.**

**The Committee endorses the government's intention for CDEI to perform a regulatory assurance role. The government should act swiftly to clarify the overall purpose of CDEI before setting it on an independent statutory footing.**

**4.6. Procurement and the Digital Marketplace**

Contributors to this review emphasised the importance of 'ethics by design'. Some ethical requirements will require technical solutions, which will need to be specified in the commissioning and design of any project. For example, to build an AI system that is accountable, public bodies may need to 'build in' the capacity for it to produce explanations for its decisions. This makes procurement a crucial point in the AI lifecycle where provisions for ethical standards must be set. It is important from the start of any project that the business, technology and procurement are aligned around what the preferred outcomes will be.

Evidence gathered for this review indicates that most public bodies will use external suppliers to build and manage their AI systems. This raises additional issues over and above those where AI is built and managed in-house. In its 2014 report, 'Ethical Standards for Providers of Public Services' and later in its 2018 report 'The Continuing Importance of Ethical Standards for Public Service Providers', the Committee called for public service providers to recognise that the Nolan Principles apply to them. Private providers of public services cannot delegate responsibility for standards to public bodies, and they should have in place provisions for ensuring high ethical standards in public service delivery, irrespective of whether they are using AI. Government also has a responsibility to manage third-party contracts in a way that engenders high ethical standards. Conversely public authorities cannot outsource their risk to suppliers.



“The Cabinet Office should reinforce the message that the Seven Principles of Public Life apply to any organisation delivering public services.

The Cabinet Office should ensure that ethical standards reflecting the Seven Principles of Public Life are addressed in contractual arrangements, with providers required to undertake that they have the structures and arrangements in place to support this.

Commissioners of services should include a Statement of Intent as part of the commissioning process or alongside contracts where they are extended, setting out the ethical behaviours expected by government of the service providers.”<sup>70</sup>

### **Recommendations from the Committee’s 2014 and 2018 reports into providers of public services**

In its 2018 report, the Committee found that the government had made some improvements in how it manages the ethical conduct of contractors as part of a broader maturing of outsourcing practices. However, the Committee has not had a formal response from government to the recommendations made in that report, and there appears to be limited progress on introducing formal measures to reinforce the application of ethical standards in the procurement process.

This lack of focus on ethical standards was reflected in evidence collected for this review. Ethical considerations do not appear to play much part in AI procurement across the public sector at present. Public policy officials and private service providers both told the Committee that provisions for ethics are not typically part of tenders or contracts, and that ethics are often considered, if at all, mid-way through the development of an AI system.

“Ethical standards are definitely not part of the procurement process at this point in time.”

**Ian O’Gara, Accenture**

Ethical considerations need to be injected early into the procurement cycle to give them the best chance of surviving the life of the contract. Ethics should be considered at each stage of the procurement process: from strategic planning, through scrutinising tenders and verifying contracts, to monitoring and evaluating the performance of a public service provider.

The procurement process should be designed so that AI products and services that facilitate high standards are preferred, and that it rewards companies that have prioritised ethical practices. As part of the commissioning process, or when contracts are extended, the government should set out the ethical principles that companies providing services to them are expected to exemplify. Adherence to ethical standards should be part of the evaluation process and should be given an appropriate weighting. Companies that show a commitment to these ethical behaviours should be scored more highly than those that do not. This would help ensure that the suppliers who think about ethics, and who build ethics into their systems, have a competitive advantage. In doing so, government will leverage its significant purchasing power to incentivise private providers to build ethical AI.

Several contributors to this review also suggested that public bodies should ensure that provisions for ethical standards are written into service delivery contracts. This was seen as particularly important given the potential for private companies to cite commercial confidentiality or trade secrets as reasons to withhold information about how their algorithms work. This would undermine

70 The Committee on Standards in Public Life (2018), The Continuing Importance of Ethical Standards for Public Service Providers. Available at: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/705884/20180510\\_PSP2\\_Final\\_PDF.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/705884/20180510_PSP2_Final_PDF.pdf)



accountability by making access to explanations and the auditing of AI systems more difficult.

“Assertions of commercial confidentiality should not be accepted as an insurmountable barrier to appropriate rights of access to the [algorithmic] 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 standard.”<sup>71</sup>

**Marion Oswald, Senior Fellow in Law and Director of the Centre for Information Rights, University of Winchester**

### **Recommendation 5:**

**Government should use its purchasing power in the market to set procurement requirements that ensure that private companies developing AI solutions for the public sector appropriately address public standards.**

**This should be achieved by ensuring provisions for ethical standards are considered early in the procurement process and explicitly written into tenders and contractual arrangements.**

Centralised procurement tools should also be improved. The Cabinet Office informed the Committee that the Crown Commercial Service’s Digital Marketplace is responsible for around 25% of public sector technology procurement of common goods and services. As it stands, the marketplace contains no provisions to support ethical standards. In order to advertise their products or services on the marketplace, private companies need only to fill out a tick-box questionnaire, with no reference to managing standards.

This represents a missed opportunity. The marketplace could offer a range of tools to help providers assess if AI products will support or undermine public standards. Canada, for example, operates a register of responsible AI companies.<sup>72</sup> The marketplace could also allow AI products and services to be classified according to certain features, such as explainability. Such tools would help public bodies navigate the range of products and services offered. In discussions with the Committee, Crown Commercial Service (CCS) officials expressed a desire for the marketplace to play a more active role in the procurement process.

A new specialist AI framework, including separate streams for machine learning and robotic process automation, is currently under development. Before the launch of its new AI framework, CCS should consider what tools it can introduce to the marketplace to best help public bodies find AI products and services that meet their ethical requirements. The shaping of this will be determined by engaging with the market, both suppliers and departments, to get their views prior to designing the new commercial vehicle.

71 Written evidence 4 (Marion Oswald)

72 Available at: <https://www.canada.ca/en/government/system/digital-government/modern-emerging-technologies/responsible-use-ai/list-interested-artificial-intelligence-ai-suppliers.html>



### **Recommendation 6:**

**The Crown Commercial Service should introduce practical tools as part of its new AI framework that help public bodies, and those delivering services to the public, find AI products and services that meet their ethical requirements.**

#### **4.7. Impact assessment**

Contributors cited the absence of a compulsory standards risk management tool as a major gap in the UK's national AI governance framework. Multiple public policy experts told this review that mandatory impact assessments should fill this gap.

This was for four reasons. First, mismanagement of AI systems could seriously undermine public standards. Impact assessments would inform public bodies what level of risk their AI system could pose, and allow those authorities to set risk-based governance accordingly.

*“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.”<sup>73</sup>*

**Centre for Data Ethics and Innovation**

Second, impact assessments were deemed necessary because AI is new technology that most public bodies have little experience in. Risks such as data bias would be new and unfamiliar to most, and so impact assessment would push these issues to the fore.

Third, impact assessments were seen as important for accountability. Proper accountability depends on public bodies being aware of the risks of their AI systems, so that authorities can be assessed against any mitigation measures they take.

Fourth, impact assessments are necessary because AI tools are likely to have a major impact on citizens and we need to be certain that their interests and rights are protected. Impact assessment is one major element in meeting the responsibility of due diligence.

Though some standards issues are covered by EIAs (Equality Impact Assessments) and DPIAs (Data Protection Impact Assessments), neither was seen as comprehensively covering all relevant standards issues. In particular, contributors cautioned against using DPIAs as a proxy for ethical risk assessment.

In contrast, multiple contributors spoke favourably about the Canadian model of Algorithmic Impact Assessment. Following a Treasury Directive on Automated Decision Making,<sup>74</sup> the Canadian government introduced a mandatory algorithmic impact assessment for automated decision systems.<sup>75</sup> The assessment consists of an electronic survey that covers the social, environmental, and human rights impact of an AI system, as well as provisions for data quality and human responsibility. It then generates a risk score for the automated system. Though some contributors noted flaws with the specific wording of the Canadian model, it was applauded as an overall framework for upholding ethical standards.

73 Written evidence 18 (Centre for Data Ethics and Innovation)

74 Available at: <https://docs.google.com/document/d/1LdcIG-UYeokx3U7ZzRng3u4T3IHrBXXk9JddjueQok/edit>

75 Available at: <https://open.canada.ca/aia-eia-js/?lang=en>



“The AIA provides designers with a measure to evaluate AI solutions from an ethical and human perspective, so that they are built in a responsible and transparent way. For example, the AIA can ensure economic interests are balanced against environmental sustainability.

The AIA also includes ways to measure potential impacts to the public, and outlines appropriate courses of action, like behavioral monitoring and algorithm assessments.”<sup>76</sup>

#### **Canadian Government Video on AIA**

Will you have documented processes in place to test datasets against biases and other unexpected outcomes? This could include experience in applying frameworks, methods, guidelines or other assessment tools.

Will you be developing a process to document how data quality issues were resolved during the design process?

Will you be making this information publicly available?

Will you undertake a Gender Based Analysis Plus of the data?<sup>77</sup>

#### **Questions on data quality taken from Canada’s Algorithmic Impact Assessment**

Alternatively, the section in the Turing AI Guide ‘Using AI ethically and safely’ favoured a Stakeholder Impact Assessment (SIA). An SIA encourages public bodies to identify affected stakeholders, analyse the fairness of desired outcomes, and examine the possible impacts of an AI system on the individual and society.

In contrast with the Canadian box-ticking approach, the SIA offers more open-ended questions and allows public bodies to develop their own sector-specific and use case-specific questions.

Both the AIA and the SIA would help public bodies navigate the full range of ethical risks that AI poses. The SIA in particular is designed to point towards effective mitigation measures. The stipulation that an SIA is carried out at three stages of an AI project lifecycle – problem formulation, pre-implementation and reassessment after deployment – ensures that its outcomes will inform project design and lead to remedial action.

#### **Goal-Setting and Objective-Mapping**

How are you defining the outcome (the target variable) that the system is optimising for? Is this a fair, reasonable, and widely acceptable definition? Does the target variable (or its measurable proxy) reflect a reasonable and justifiable translation of the project’s objective into the statistical frame?

Is this translation justifiable given the general purpose of the project and the potential impacts that the outcomes of its implementation will have on the communities involved?<sup>78</sup>

#### **Questions taken from the UK government guidance’s Stakeholder Impact Assessment**

76 Available at: <https://www.canada.ca/en/government/system/digital-government/modern-emerging-technologies/responsible-use-ai/algorithmic-impact-assessment.html>

77 Available at: <https://open.canada.ca/aia-eia-js/?lang=en>

78 Dr David Leslie (2019), Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. The Alan Turing Institute. Available at: [https://www.turing.ac.uk/sites/default/files/2019-06/understanding\\_artificial\\_intelligence\\_ethics\\_and\\_safety.pdf](https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf)



The Committee believes that for impact assessments to be effective they must meet three conditions. First, an impact assessment should be mandatory for any machine learning system before it is deployed. The Committee heard that too often in the public sector impact assessments are undertaken as a rubber-stamping activity after a project has already been approved. Given the absence of regulatory clarity, an optional impact assessment would mean that restraints on machine learning systems in the UK public sector remained weak. In line with the SIA model, the impact assessment should also be repeated at later stages as an AI system develops.

Second, as an AI impact assessment is a tool of ethical review, it should not be set by an organisation which already has a vested interest in implementing an AI system. In this regard, we do not favour the SIA's provision to allow public bodies to write their own use-case specific questions. In the words of one expert consulted, a decentralised approach would allow "departments to mark their own homework".

Third, impact assessments should be published. The Office for AI guidance advises public bodies to publish the details of their SIA pre-deployment, which would include details of any mitigation measures taken. This is vital for scrutiny and accountability, as it would allow members of the public to assess how far a public authority has managed the risk an AI system poses adequately. The government should make publication of an AI impact assessment mandatory.

Contributors noted that public sector organisations using AI would be likely to trigger the legislative requirement to undertake a data protection impact assessment, and that the creation of an extra form of impact assessment would create an additional administrative burden for public officials. Government should consider how an AI impact assessment could be integrated into the DPIA process in order to streamline this process for public sector organisations.

#### **Recommendation 7:**

**Government should consider how an AI impact assessment requirement could be integrated into existing processes to evaluate the potential effects of AI on public standards. Such assessments should be mandatory and should be published.**

#### **4.8. Transparent disclosure**

The Committee saw two public sector AI projects during the course of this review that demonstrated transparency. The West Midlands Police and Crime Commissioner's Ethics Committee, which advises on data science projects proposed by their Data Analytics Lab, publishes its minutes in full, including where they have criticised police practice.<sup>79</sup> Moorfields Eye Hospital, who have been working in partnership with DeepMind Health (now Google Health) since 2016, also have a useful section on their website dedicated to their machine learning project, including a Q&A and latest updates.<sup>80</sup>

79 Available at: <https://www.westmidlands-pcc.gov.uk/ethics-committee/>

80 Available at: <https://www.moorfields.nhs.uk/landing-page/deepmind-health-research-partnership>



Contributors working on AI projects across the public sector told the Committee that negative and sensationalist media coverage often made public bodies wary of being transparent. Policy experts also told the Committee that the biggest incentive towards transparency was often the personal ethical commitment of those working with AI in the public sector. This can lead to quite a fragmented approach to public standards. The Committee heard that central coordination around transparency is required because it is not currently mandated by any regulation or institution.

In its report on Algorithms in the Criminal Justice System, the Law Society recommended the creation of a register of algorithmic systems in criminal justice in the UK. The Committee is of the view that such a register could be expanded beyond criminal justice, if a sensible threshold is set. In discussions with the Committee, the Centre for Data Ethics and Innovation expressed an interest in overseeing such a register.

“We note the recommendation by the Law Society that a national register of automated decision making tools 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.”<sup>81</sup>

**Centre for Data Ethics and Innovation**

However, it is likely that the establishment of a central register, even if restricted to AI systems above a high threshold, would be an extensive and potentially overwhelming bureaucratic challenge, particularly given the predicted scale of AI across public life. There is no guarantee that such a register would be properly accessible to the public. Similar registers, such as those currently used to collect procurement data, were criticised by contributors to this review for being poorly formatted, incomplete and difficult to search.

Having a centralised register of AI systems would also be counter-intuitive to the general public, who would likely go to the website of a public body to find information about how they operate, rather than central government. The Committee is of the view that any system intended to increase transparency should not focus on the creation of a centralised database.

There are already requirements for proactive disclosure under the Freedom of Information Act 2000 (FOI Act). There is a statutory obligation on public bodies under Section 19 of the Act to proactively publish information that the public are likely to be interested in. It is the duty of every public authority to adopt and maintain a publication scheme, approved by the Information Commissioner, which makes information about their business activities available to the public.<sup>82</sup> As the regulator, the ICO provides public bodies with an approved model publication scheme that specifies categories of information that should be published. This includes information about income and expenditure; tendering, procurement and contracts; and decision-making processes.<sup>83</sup> Public bodies that use AI will still need to proactively disclose this information to the public. In theory, this should encourage openness and transparency.

81 Written evidence 18 (Centre for Data Ethics and Innovation)

82 Freedom of Information Act 2000, section 19. Available at: <http://www.legislation.gov.uk/ukpga/2000/36/section/19>

83 ICO [online], Freedom of Information Act: Model publication scheme. Available at: <https://ico.org.uk/media/for-organisations/documents/1153/model-publication-scheme.pdf>



The Committee heard, however, that the proactive disclosure requirements of the FOI Act have limited use in the current framework, not least because the legislation is outdated. The ICO told us that publication schemes are not necessarily useful for enforcing transparency, particularly because it is difficult to assess the nature and extent of compliance across the public sector. They said that more could be done to encourage proactive disclosure in other ways, by promoting openness and transparency by design, for example. This would require public bodies to think about what information they should proactively disclose, as well as the implications of not being transparent, from the start of the AI commissioning process.

It is unlikely, however, that an expectation on public bodies to think about openness is enough to change behaviour. Public bodies will need guidance to help them think through openness and transparency implications. The Committee recommends that government should set clear guidelines for public bodies on what information they should proactively disclose about their AI systems.

These guidelines should make explicit the features of an AI system that warrant transparency, such as the processing of personal data for predictive analytics, or the potential impact of a system on an individual. They should also specify how public organisations should make information available to the public.

### **Recommendation 8:**

**Government should establish guidelines for public bodies about the declaration and disclosure of their AI systems.**

Members of the public also have a general right of access to information under Section 1 of the FOI Act. This says that any person making a request for information to a public authority is entitled to be informed by the public authority whether it holds information of the description specified in the request, and if that is the case, to have that information communicated to them.<sup>84</sup> This does not extend to private providers of public services. The Committee has previously recommended that the government should hold a consultation on extending the application of the FOI Act to private providers where information relates to the performance of a contract with the government in the delivery of public services.<sup>85</sup> The increasing use of private sector companies to deliver AI-enabled public services adds urgency to the Committee's 2018 recommendation.

84 Freedom of Information Act 2000, section 1. Available at: <http://www.legislation.gov.uk/ukpga/2000/36/section/1>

85 The Committee on Standards in Public life (2018), *The Continuing Importance of Ethical Standards for Public Service Providers*. Available at: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/705884/20180510\\_PSP2\\_Final\\_PDF.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/705884/20180510_PSP2_Final_PDF.pdf)





# Chapter 5: The role of public bodies

## 5.1. Introduction

Decisions on adopting and implementing artificial intelligence in the public sector lie with individual government departments and public bodies. Individual police forces or NHS trusts, for example, commission and operate AI systems in their organisation. Each body will need to establish suitable governance mechanisms to manage the ethical risks associated with AI and address regulatory compliance.

In January 2019, Singapore's Personal Data Protection Commission published a proposed model framework for the governance of AI systems in Singapore.<sup>86</sup> The framework is a useful starting point for thinking about the kinds of mechanisms that public sector organisations in the UK should adopt when using AI technology. It states that the risks associated with AI can be managed by adapting existing governance structures to incorporate values, risks and responsibilities relating to algorithmic decision-making. This includes setting clear roles and responsibilities for the ethical deployment of AI and putting in place internal controls to address the risks involved in using AI to make decisions.

The Committee shares the view that effective governance of AI in the public sector does not require a radical overhaul of traditional risk management. Public sector organisations should already have in place governance frameworks that identify, assess and mitigate risk and establish clear responsibilities for decision making. They are also already subject to rigorous scrutiny and checks by external bodies to ensure that they are operating in accordance with their mandates. Therefore, it should not be a huge step for public bodies to put in place effective risk management structures to ensure the robust governance of AI.

This chapter looks first at the risks that need to be managed before deployment, when public bodies are contemplating using AI and are developing AI systems for public service delivery. It then covers five areas of governance key to risk management when deploying AI:

- setting responsibility
- internal and external oversight
- monitoring and evaluation
- appeal and redress
- training and education

Recommendations made here reflect the issues of most concern to the Committee and are intended to supplement public sector guidance discussed in chapter 3.

## 5.2. Legal and legitimate AI

Before deploying AI, public sector organisations need to demonstrate that the benefits of using the technology outweigh the risks. They will also need to ensure that they are using AI in ways that are legal and legitimate and do not undermine individual rights.

As a first step, policy experts emphasised that public bodies should carefully consider the appropriateness of using artificial intelligence in any given context. On a case by case basis, public sector organisations will need to justify why they are using an algorithm; consider whether the potential impact on individuals is necessary and proportionate; and demonstrate how the tool will improve the current system.<sup>87</sup> Office for AI guidance makes clear that the decision to use AI should always be based on user need.<sup>88</sup> Contributors to this review said that public officials should be prepared to walk away from experimental AI where

86 Personal Data Protection Commission Singapore (2019), A proposed model Artificial Intelligence governance framework (for public consultation). Available at: <https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/A-Proposed-Model-AI-Governance-Framework-January-2019.pdf>

87 Written evidence 2 (Jamie Grace)

88 Office for AI and GDS (2019), Assessing if artificial intelligence is the right solution. Available at: <https://www.gov.uk/guidance/assessing-if-artificial-intelligence-is-the-right-solution>



there is no clear benefit to the public and where the potential infringement of individual rights cannot be shown to be necessary and proportionate.<sup>89</sup> The Committee agrees with this judgement.

*“You can imagine a scenario where things go wrong because the public sector has implemented some AI technology because it is shiny, cool and exciting rather than helpful.”*  
**Eddie Copeland, Director, London Office of Technology and Innovation (LOTI)**

All providers of public services must also publish a statement on how their use of AI complies with relevant laws and regulations, as the Committee recommends in chapter 4 (recommendation 2). This is of particular importance given that these technologies can interfere with individual rights and freedoms, and do so at scale, operating in ways that are often difficult for individuals to understand, challenge or contest.

Where AI automates the processing of personal data, public bodies will also need to demonstrate that the data processed by the algorithm is fairly and lawfully obtained, processed and retained, and only used for legitimate purposes, as stipulated under the GDPR. These issues will probably need to be considered through some form of data protection and/or AI impact assessment (as discussed in chapter 4).

### 5.3. System design

Experts consulted for this review said that good system design could help to mitigate some of the risks to standards identified in this report. For example, it may be possible to build into a system a degree of technical transparency,

or provisions for monitoring and evaluating an AI system’s performance. To do this, public bodies will need to anticipate how public standards may be affected by any new system before it is introduced and subsequently in deployment.

Public bodies should start by conducting an AI impact assessment, as discussed in chapter 4. This will help public organisations assess how their proposed AI system could affect public standards such as openness, accountability and objectivity. During the course of this review, some public officials expressed concerns that impact assessments could be used retrospectively after the details of a project were already set, as a tick box exercise to show compliance rather than the proper consideration of ethical risk. Such an approach should be avoided as it could embed avoidable risk into the design of an AI system.

Where standards risks are identified it is essential that project development teams alter the design of their systems. For example, if a project risks amplifying bias, public bodies may want to consider broadening their dataset to dilute the effects of that bias. Similarly, if a system is highly automated and risks undermining the principle of accountability, public bodies should consider redesigning the system so that human involvement in the decision-making process is active and meaningful. There is no one-size-fits-all answer to AI system design, as this will be highly context and risk dependent. In the case of automation and accountability, for example, it may be acceptable to automate a system that sends citizens text reminders to pay their council tax, but it would not be appropriate to automate a predictive policing system that grants or denies parole to prisoners.

89 Written evidence 4 (Marion Oswald)



Contributors also emphasised that the type of standards risk identified at the project design stage should inform decisions as to whether to procure an AI system from external providers or to build one in-house. One expert cited West Midlands Police as an example of good practice: by having their own in-house Data Lab, developers understand the ethical constraints of a policing context and apply that understanding when designing their AI systems.

If the risk to public standards remains high despite any mitigation measures taken, then public bodies should not shy away from moderating or constraining the intended use of an AI system. In some cases, it may be that a project should not proceed from design to deployment, even if significant expenses have already been incurred. One expert said that an AI system could come with a “health warning” if there was a high risk the product was biased because it had only been trained on certain populations.

Finally, contributors emphasised the importance of considering standards iteratively as a project progresses from design through to deployment. This is because some systems will undergo a process of continuous testing and redesign to optimise performance, and because AI systems can act in ways that are unpredictable and unexpected. The values underlying what is acceptable and unacceptable in different contexts may also change over time.

Public bodies should not only mitigate standards risks at the project design stage, but continue to monitor risks to standards at all stages of the AI lifecycle and throughout the duration of an AI project. Standards review will need to occur every time a substantial change to the design of an AI system is made.

### **Recommendation 9:**

**Providers of public services, both public and private, should assess the potential impact of a proposed AI system on public standards at project design stage, and ensure that the design of the system mitigates any standards risks identified.**

**Standards review will need to occur every time a substantial change to the design of an AI system is made.**

### **5.4. Diversity**

The field of AI is at risk of replicating or perpetuating historical biases and existing structures of inequality in society (see 2.5). In April 2019, a report by the AI Now Institute said that biased AI systems can largely be attributed to the lack of diversity within the AI industry.<sup>90</sup>

Public bodies must maximise diversity at all stages of the AI process to help tackle issues of bias and discrimination within AI systems. There needs to be diversity in the workforce and in training and education, so that biases, whether conscious or unconscious, are less likely to be programmed into AI systems. This includes those building and developing AI systems, and those who have responsibility for AI at various stages of deployment (see 5.5.1). An increased access to a wider range of skills and perspectives at each stage of the process will help public bodies to better consider the impact of AI systems on public standards, and to mitigate the risks identified. Datasets used to train machine learning algorithms will also need to be diverse, so that they work accurately and objectively on different individuals and populations.

90 Sarah Myers West, Meredith Whittaker and Kate Crawford (2019), ‘Discriminating Systems: Gender, Race and Power in AI’, AI Now Institute. Available at: <https://ainowinstitute.org/discriminatingystems.pdf>



**Recommendation 10:**

**Providers of public services, both public and private, must consciously tackle issues of bias and discrimination by ensuring they have taken into account a diverse range of behaviours, backgrounds and points of view. They must take into account the full range of diversity of the population and provide a fair and effective service.**

The Committee has identified five areas of governance necessary for upholding public standards in this context: setting responsibility; monitoring and evaluation; internal and external oversight; appeal and redress; and training and education.

**5.5.1. Setting responsibility**

In most AI systems, there will not be a single person responsible for the whole system; rather responsibility will be allocated across a range of individuals who engage with the system at various stages of deployment. The key question for public bodies is how responsibility for, and oversight of, AI is allocated across an organisation.

**5.5. Deployment of an AI system**

Once public sector organisations have assessed their proposed AI systems to show that they are necessary, proportionate and lawful, and have designed their systems in ways that help to mitigate the ethical risks identified, they need to set effective governance mechanisms for its use. Even well designed AI systems will pose risks to openness, accountability and objectivity, and so public bodies will need to put in place sound risk management and other internal controls to address those risks in the day-to-day management of the system.

Responsibility, and ultimately human control, will be shared by individuals from across an organisation, including individuals who operate AI systems, project managers who monitor entire AI systems and senior leadership who oversee the policy for which AI is being used. Responsibility could be distributed as in the table below.

Senior leadership	Project managers	Individuals operating AI systems
Make decision to introduce an AI system	Oversee end-to-end AI system process	Check input data
Set governance mechanisms for AI system	Assess the impact of the AI on groups of data subjects	Identify any false positives or system errors
Assess how the AI impacts their policy area as a whole	Monitor and evaluate the system, and make improvements where necessary	Accept or reject decision recommendations



As the Office for AI set out in their guidance, any allocation of responsibility should be clearly documented, so public officials are fully aware of their roles and responsibilities and it is clear to all officials interacting with an AI system where responsibility lies. Such a record will help facilitate accountability for the system.

“Humans must be ultimately responsible for decisions made by any system...Good governance will require for each use case, a specific understanding of the appropriate division of responsibilities.”<sup>91</sup>

**Centre for Data Ethics and Innovation**

Once responsibility is set, senior leadership must ensure that public officials have the capacity to exercise their responsibility in a meaningful way. They must be properly trained and provided with the resources and guidance needed for them to discharge their duties. Fundamentally, officials must have both the knowledge and the power to implement change, otherwise any designated responsibility is meaningless.

“The person [needs to have] both the agency and the knowledge necessary to make changes to the system’s behaviour and to intervene when it seems like something is going to go wrong.”

**Dr Brent Mittelstadt, Research Fellow and British Academy Postdoctoral Fellow, Oxford Internet Institute**

Public bodies should also seek to ensure that their organisational culture encourages and empowers their officials to use their professional knowledge and expertise in confirming automated decisions.

Similarly, private sector providers should make sure that any system they build meets the requirements for human responsibility set by the public authority.

### **Recommendation 11:**

**Providers of public services, both public and private, should ensure that responsibility for AI systems is clearly allocated and documented, and that operators of AI systems are able to exercise their responsibility in a meaningful way.**

#### **5.5.2. Monitoring and evaluation**

Public bodies deploying AI should establish monitoring systems and processes to evaluate and identify issues relating to the performance of the technology. It is not acceptable for a public organisation introducing AI to assume the technology will always function as intended, particularly because machine learning systems are often vulnerable to flaws like inaccuracy, and can operate in unique and unexpected ways that can have unintended consequences.

Some contributors to this review argued that public bodies may not be aware AI systems could be inaccurate, often citing facial recognition as an example. At Notting Hill Carnival in 2017, facial recognition technology used by the Metropolitan Police was said to be wrong 98% of the time, and more likely to misidentify ethnic minorities and women.<sup>92</sup> The possibility of inaccuracy underscores the importance of monitoring and evaluation, particularly in a context like facial recognition, where the consequences of misidentifying an individual can be significant. Public bodies cannot assume that their AI systems will work as well in real life as they

91 Written evidence 18 (Centre for Data Ethics and Innovation)

92 Vikram Dodd (2018), ‘UK police use of facial recognition technology a failure, says report’, The Guardian. Available at: <https://www.theguardian.com/uk-news/2018/may/15/uk-police-use-of-facial-recognition-technology-failure>



do in a data science lab. They will also need to judge what an acceptable error rate is for their specific use of AI, according to the probability and severity of harm to an individual in a particular context.

Contributors also told the Committee that monitoring is vital to prevent unintended consequences. Even where AI is introduced with good intentions, poor quality data or a lack of knowledge about how an AI system operates will lead to unwanted outcomes. Public bodies should periodically re-test and validate their models on different demographic groups to observe whether any groups are being systematically advantaged or disadvantaged, so that they can update their AI systems where necessary.

Many machine learning systems also refine the way that they process data to improve accuracy over time. Such refinements may distort the original goal of the AI system, so public bodies will need to monitor whether an AI system is achieving its intended purpose. The continuous refinement of AI systems could also be a problem if the system is deployed in an environment where the user can alter its performance and does so maliciously. For example, Microsoft's chatbot Tay was designed to learn from interactions it had with real people on Twitter in 2016. When users decided to feed it racist and offensive information, it learned to interact that way itself.<sup>93</sup> Chatbots that dispense advice on behalf of public bodies will have significant effects on citizens. These systems will need to be subject to consistent monitoring and evaluation to ensure they are not corrupted by human interaction, either intentionally or by accident.

“Another concern is when you have systems that continue to learn through interaction with the user. There is the potential for a user to either maliciously poison the training data or to be mischievous in the way that they train the system thereby influencing the way it develops in the future.”

**Fiona Butcher, Fellow, Defence, Science and Technology Laboratory, Ministry of Defence**

Deploying AI is a process, not a single event. Once established, public bodies will need to keep a close eye on their AI systems to ensure that they continue to operate as intended. Deriving the best outcomes from AI will require a continuous process of tweaking and moderating the way an AI system operates.

#### **Recommendation 12:**

**Providers of public services, both public and private, should monitor and evaluate their AI systems to ensure they always operate as intended.**

93 Dave Lee (2016), 'Tay: Microsoft issues apology over racist chatbot fiasco', BBC News. Available at: <https://www.bbc.co.uk/news/technology-35902104>



### 5.5.3. Internal and external oversight

Human oversight of AI is a standards imperative. To ensure that public bodies remain accountable for automated decision-making, there needs to be internal control over the AI system, its decision-making process and its outcomes. Senior leadership should oversee the entire end-to-end AI process, to ensure that potential violations of human rights and public standards do not occur. While monitoring and evaluation requires a detailed look at data input and output, proper oversight involves leadership taking a bird's eye view of an entire system, including its design, governance and outcomes.

To have complete control over their AI systems, senior leadership need to have oversight over the whole AI process, from the point of data entry to the implementation of an AI-assisted decision.

Currently, a number of civil society organisations exercise, at a distance, oversight of AI systems. Organisations such as Liberty and Big Brother Watch have been prominent in scrutinising live facial recognition and predictive policing technologies. While these organisations are a vital part of democratic accountability, some contributors expressed doubts that civil society alone can provide meaningful oversight when AI is deployed across government at scale. Other external oversight mechanisms, such as ethics committees, will probably be necessary, as will good regulation, as discussed in chapter 4.

“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.”<sup>94</sup>

#### **Law Society Report, Algorithms in the Criminal Justice System**

Oversight is important for proper accountability as the perspectives of those running AI systems on a day-to-day basis might be quite limited. For example, an individual whose role it is to decide whether to accept or reject an AI decision may not be aware of how data input has affected the outcome. Similarly, those building datasets to feed into an algorithm may not be aware of how their input selections could adversely affect decisions made later on. Effective oversight will help public bodies identify misuse and other unintended consequences of AI.

Specialist oversight bodies are useful tools for ensuring that difficult ethical issues relating to AI are given proper consideration. In high-risk areas such as health, policing or criminal justice, the use of an independent ethics committee would help ensure that issues around openness, accountability and objectivity are considered by individuals with the necessary knowledge and expertise. They also provide an independent level of assurance and are less likely to be subject to conflicts of interest. Some contributors to the review also noted that formal ethics committees could help build public trust for new technologies.

94 The Law Society (2019), Algorithms in the Criminal Justice System. Available at: <https://www.lawsociety.org.uk/support-services/research-trends/algorithm-use-in-the-criminal-justice-system-report/>



“We use oversight bodies to assure ourselves that we have consent from the public because we know that the people who are most likely to be adversely affected by AI are less likely to come forward and present their views. We use oversight bodies, scrutiny panels and independent advisory groups to be representative of those communities.”

**Superintendent Chris Todd,  
West Midlands Police**

Public bodies need to choose oversight mechanisms that are appropriate for the systems they are developing. If the risk to individuals is low, internal oversight of AI by senior leadership may be sufficient. In other, more risky policy areas, external scrutiny may be necessary. One contributor suggested that public bodies could appoint an independent ethics officer to oversee these considerations.

Contributors also suggested that public bodies should be required to act on recommendations made by independent oversight bodies, so that they had real powers of scrutiny. The Committee agrees, and is of the view that public bodies should act on any recommendations made by independent oversight bodies and set oversight mechanisms that allow their AI systems to be properly scrutinised.

### **Recommendation 13:**

**Providers of public services, both public and private, should set oversight mechanisms that allow for their AI systems to be properly scrutinised.**

#### **5.5.4. Appeal and redress**

To remain accountable for their decisions, public bodies need to enable people to challenge decisions and to seek redress using procedures that are independent and transparent. This is the case whether AI is involved in the decision-making process or not. This is because public bodies and organisations carrying out public functions have to act in accordance with public and administrative law principles, and must act lawfully, rationally, proportionately and fairly.<sup>95</sup> Public law allows citizens to contribute to a public body’s decision-making process, through consultation, and to challenge individual decisions where they have been made.

Many public bodies have complaints procedures that individuals can follow. Where complaints cannot be resolved, individuals usually have access to independent and impartial advice through an ombudsman scheme, and almost all decisions made by public bodies that have an impact on citizens carry a statutory right of appeal. This means that decisions will need to be explained and justified to a tribunal or other independent body, irrespective of whether AI is used. Appeals can generally look at whether the decision was made in accordance with the law and make findings of fact. Individuals can also ask public bodies to review their decisions in certain circumstances. If there is no right of appeal, complaints procedure, ombudsman scheme or review process – or if those things do not adequately address the problem – individuals may be able to challenge a decision by judicial review.

95 Public Law Project (2018), An introduction to public law. Available at: <https://publiclawproject.org.uk/wp-content/uploads/2018/07/An-introduction-to-public-law-1.pdf>





The existing appeals process should be utilised for those wishing to appeal against automated decisions. Public bodies should continue to make available fair and transparent avenues of redress for individuals who have been adversely affected by a decision, even if that decision was automated. Whether AI is used or not, public bodies should continue to uphold existing principles of administrative justice. For example, public bodies should: (1) make users and their needs central when making decisions; (2) enable people to challenge decisions and seek redress in ways that are independent, fair and transparent; and (3) keep people informed and empower them to resolve their problems as quickly and comprehensively as possible.<sup>96</sup> Public bodies should also ensure that their mechanisms for redress continue to be proportionate and efficient, and lead to well-reasoned, lawful and timely outcomes.

Public bodies will need to be able to explain and justify decisions made by AI technology. This means that they need to be auditable and transparent enough to satisfy a proper process of appeal and redress. Audits are necessary to discover how AI systems work and make decisions. Public bodies need to be able to track the process by which a system was designed, procured and deployed, and should be able to trace the way an automated decision was made. A decision that adversely affects an individual may be down to the failure of any one of these stages or a combination of them. A meaningful process of redress should enable public bodies to find out what failed and how that failure can be rectified.

#### **Recommendation 14:**

**Providers of public services, both public and private, must always inform citizens of their right and method of appeal against automated and AI-assisted decisions.**

#### **5.5.5. Training and education**

Contributors to this review consistently emphasised the importance of training and education. Public officials will need new skills and knowledge to ensure that high public standards are upheld in an AI-enabled public sector. Those using AI at all levels will need to be taught how AI works and be educated about the ethical risks of AI systems.

Contributors told this review that the risks to standards were greater if the decision to introduce AI was poorly informed. Without the right knowledge, senior management may be unaware, for example, of the potential for AI to amplify historic bias in their policy area.

#### **Working with the right skills to assess AI**

When identifying whether AI is the right solution, it's important that you work with:

- specialists who have a good knowledge of your data and the problem you're trying to solve, such as data scientists
- at least one domain knowledge expert who knows the environment where you will be deploying the AI model results.<sup>97</sup>

**Office for AI Guidance, Assessing if artificial intelligence is the right solution**

96 Administrative Justice and Tribunals Council (2010), Principles for Administrative Justice

97 Government Digital Service and Office for AI (2019), Assessing if artificial intelligence is the right solution. Available at: <https://www.gov.uk/guidance/assessing-if-artificial-intelligence-is-the-right-solution>



Those commissioning public services will need to know the technical capabilities of AI systems to assess the risks to standards posed by AI. Commissioners of public services will require a reasonable level of technical knowledge to judge whether the level of explainability a system offers matches the need for explanations in their policy area, for example.

Those operating AI systems will need to understand how their precise system operates to identify errors, ensure data is input correctly, and exercise discretion when implementing AI-enabled decisions. If operators of AI systems are not suitably trained, it would be unreasonable to hold them accountable for accepting or rejecting an AI decision.

“From the perspective of the judiciary or the courts, I think education is the starting point... we are going to have to do a lot of work to develop effective training, knowledge systems and skills systems, to enable judges as well the Court Service staff to understand the implications of the operations of the systems.”  
**John Sorabji, Principal Legal Adviser to the Lord Chief Justice and Master of the Rolls**

Training and education should happen before an AI system is deployed, but it should not be a one-off event. AI experts told this review that, like any new technology, AI is still in a period of rapid change. Individual systems themselves would be continuously upgraded and their capabilities enhanced. Training and education will have to keep up with these changes. It is important, therefore, that training and education is an ongoing process throughout the lifecycle of an AI system and not a one-off event.

### **Recommendation 15:**

**Providers of public services, both public and private, should ensure their employees working with AI systems undergo continuous training and education.**

Citizens will also need to be informed about how artificial intelligence will change the way they engage with public services. Government should publicise information on citizens’ data rights and facilitate better public understanding of how AI-enabled public services will operate. Public engagement on AI will help increase trust in government innovation and ensure citizens do not feel disempowered by new technology. It is part of the CDEI’s remit to lead on public engagement and the Committee believes this should be a vital part of the Centre’s role in public life.



# Appendix 1: About the Committee on Standards in Public Life

The Committee on Standards in Public Life (CSPL, the Committee) advises the Prime Minister on ethical standards across the whole of public life in England. It monitors and reports on issues relating to the standards of conduct of all public office-holders. The Committee is an advisory non-departmental public body sponsored by the Cabinet Office. The Chair and members are appointed by the Prime Minister.

The Committee was established in October 1994, by the then Prime Minister, with the following terms of reference:

“To examine current concerns about standards of conduct of all holders of public office, including arrangements relating to financial and commercial activities, and make recommendations as to any changes in present arrangements which might be required to ensure the highest standards of propriety in public life.”

The remit of the Committee excludes investigation of individual allegations of misconduct.

On 12 November 1997, the terms of reference were extended by the then Prime Minister: “To review issues in relation to the funding of political parties, and to make recommendations as to any changes in present arrangements.”

The terms of reference were clarified following the Triennial Review of the Committee in 2013. The then Minister for the Cabinet Office confirmed that the Committee “should not inquire into matters relating to the devolved legislatures and governments except with the agreement of those bodies”, and that “the Government understands the Committee’s remit to examine ‘standards of conduct of all holders of public office’ as encompassing all those involved in the delivery of public services, not solely those appointed or elected to public office.”

The Committee is a standing committee. It not only conducts inquiries into areas of concern about standards in public life, but can also revisit those areas to monitor whether and how well its recommendations have been put into effect.

## **Membership of the Committee for the period of this review**

Lord (Jonathan) Evans KCB DL, Chair  
The Rt Hon Dame Margaret Beckett DBE MP  
The Rt Hon Jeremy Wright QC MP (from 21 November 2019)  
The Rt Hon Simon Hart MP (until July 2019)  
Dr Jane Martin CBE  
Jane Ramsey  
Dame Shirley Pearce DBE  
Monisha Shah  
The Rt Hon Lord (Andrew) Stunell OBE

## **Chair of Committee’s Research Advisory Board**

Professor Mark Philp

## **Secretariat**

The Committee is assisted by a Secretariat consisting of Lesley Bainsfair (Secretary to the Committee), Amy Austin (Senior Policy Adviser), Ally Foat (Senior Policy Adviser), Nicola Richardson (Senior Policy Adviser), Aaron Simons (Senior Policy Adviser) and Lesley Glanz (Executive Assistant). Press support is provided by Maggie O’Boyle.



# Appendix 2: Terms of reference

The terms of reference for the Committee's review into artificial intelligence and standards are 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, including:
  - a. examining the current use of artificial intelligence and associated advanced technologies in the public sector
  - b. exploring how standards may be affected by the widespread introduction of these technologies into the public sector
  - c. examining what safeguards and considerations of standards are currently in place in technology procurement processes in the public sector
  - d. examining what safeguards and considerations of standards are currently in place in the deployment of AI and advanced technologies within the public sector
  - e. examining what safeguards and considerations of standards are currently in place in private sector organisations developing AI services intended for use 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.

# Appendix 3: Methodology

The Committee used a range of methods as part of its evidence gathering for its review, including:

- 50 individual stakeholder meetings and conference calls
- 3 roundtable seminars
- 19 written submissions
- polling and focus group research
- desk research, including a review of relevant academic texts, think tank reports, government and parliamentary reviews, and media coverage
- attending AI roundtables and conferences hosted by external organisations.

## Stakeholder meetings

The Committee and Secretariat held 50 meetings and conference calls with individual stakeholders.

Name	Organisation
Carly Kind and Olivia Varley-Winter	Ada Lovelace Institute
Professor Edward Harcourt	Arts and Humanities Research Council (AHRC)
Emily Commander	Arts and Humanities Research Council (AHRC)
Tabitha Goldstaub	AI Council
Dr Adrian Weller, Dr David Leslie, Dr Florian Ostmann and Dr Ricardo Silva, Dr Brent Mittelstadt	The Alan Turing Institute
Dr David Halpern and Aisling Ní Chonaire	Behavioural Insights Team
Silkie Carlo	Big Brother Watch
Gillian Stamp	Bioss International
Crofton Black	The Bureau of Investigative Journalism
Roger Taylor, Alex Lawrence-Archer, Oliver Buckley, Bethan Charnley and Michael Birtwistle	Centre for Data Ethics and Innovation (CDEI)
Richard Thomas CBE and Bojana Bellamy	Centre for Information Policy Leadership
Chief Rabbi Ephraim Mirvis	Chief Rabbi of the United Hebrew Congregations of the Commonwealth
Niall Quinn	Crown Commercial Service
Joe Baddeley, Sam Roberts and Natalia Domagala	Department for Digital, Culture, Media and Sport (DCMS)
Rebecca Hilsenrath and Andrew Harding	Equality and Human Rights Commission (EHRC)
Steve Unger	Flint Global
Jacob Turner	Fountain Court Chambers
Matthew Cain, Robert Miller, Liz Harrison and Suki Binjal	Hackney Council
Apollo Gerolymbos	London Fire Brigade



Appendix 3: Methodology

Name	Organisation
Christophe Prince	Home Office
Lord Clement-Jones	House of Lords APPG on AI
Professor Nick Jennings	Imperial College London
Simon McDougall	Information Commissioner's Office
Chief Constable Alan Pughsley	Kent Police
Christina Blacklaws and Alexandra Cardenas	The Law Society
Dr Rune Nyrup	Leverhulme Centre for the Future of Intelligence, University of Cambridge
Eddie Copeland	London Office of Technology (LOTI)
Professor Richard Susskind	IT Adviser to the Lord Chief Justice
Dr Pearse Keane	Moorfields Eye Hospital
Sarah Wilkinson	NHS Digital
Matthew Gould	NHSX
Jacob Beswick, Tim Cook and Sabine Gerdon	Office for AI
Professor Sandra Wachter	Oxford Internet Institute, University of Oxford
Dr Jonathan Bright	Oxford Internet Institute, University of Oxford
Professor Luciano Floridi	Oxford Internet Institute, University of Oxford
James Loft and Nisha Deo	Rainbird Technologies
Alexander Babuta	Royal United Service Institute (RUSI)
Simon Dennis	SAS Institute
Zee Kin Yeong	Singapore Infocomm Media Development Authority
Ed Humpherson	UK Statistics Authority
Professor Alastair Denniston	University of Birmingham
Professor Karen Yeung and Professor Andrew Howes	University of Birmingham
Professor Charles Raab	University of Edinburgh
Professor Dame Wendy Hall	University of Southampton
Superintendent Iain Donnelly	West Midlands Police
Thomas McNeil	Strategic Adviser to the West Midlands Police and Crime Commissioner



## Roundtable seminars

The Committee held three roundtable seminars in London as part of this review. Transcripts of the roundtables are available on the Committee's website.

Roundtable for Practitioners, Government and Public Service Providers, held on 23 May 2019, at 1 Horse Guards Road, London

Lord (Jonathan) Evans	Chair, Committee on Standards in Public Life (CSPL)
Dame Shirley Pearce	Independent Member, Committee on Standards in Public Life (CSPL)
Jane Ramsey	Independent Member, Committee on Standards in Public Life (CSPL)
Monisha Shah	Independent Member, Committee on Standards in Public Life (CSPL)
Oliver Buckley	Executive Director, Centre for Data Ethics and Innovation (CDEI)
Fiona Butcher	Science and Technology Lab, Ministry of Defence
Bethan Charnley	Innovation Policy Lead, Government Digital Service
Jimmy Elliott	General Counsel, SAS Institute
Sabine Gerdon	Project Lead, Office for AI
Sana Khareghani	Head of the Office for AI
Alex Lawrence-Archer	Chief Operating Officer, CDEI
Simon McDougall	Executive Director for Tech, Policy and Innovation, ICO
Ian O'Gara	Digital Strategy Director (Public Sector), Accenture
Marion Oswald	Senior Fellow, Department of Law, University of Winchester



Roundtable for Academics and Policy Experts,  
held on 29 May 2019, at Imperial College London.

Lord (Jonathan) Evans	Chair, Committee on Standards in Public Life (CSPL)
Jane Ramsey	Independent Member, Committee on Standards in Public Life (CSPL)
Monisha Shah	Independent Member, Committee on Standards in Public Life (CSPL)
Professor Mark Philp	Chair, Research Advisory Board, Committee on Standards in Public Life (CSPL)
Professor Nick Jennings	Vice Provost, Imperial College
Alexander Babuta	Research Fellow, National Security Studies, RUSI
Professor Alan Brown	Professor in Digital Economy, University of Exeter
Alexandra Cardenas	Head of Commercial and Technology Law, The Law Society
Jamie Grace	Senior Lecturer in Law, Sheffield Hallam University
Professor Edward Harcourt	Director of Research, AHRC, UKRI
Professor Philip Howard	Director, Oxford Internet Institute
Samantha McGregor	Head of Creative Industries, Digital Arts and Humanities, AHRC, UKRI
Professor Charles Raab	Professorial Fellow, University of Edinburgh
Chief Superintendent Chris Todd	NPCC lead for Data Analytics, and Director of Intelligence at West Midlands Police
Peter Wells	Director of Public Policy, Open Data Institute





Roundtable for Academics and Policy Experts,  
held on 5 June 2019, at Admiralty House, London

Lord (Jonathan) Evans	Chair, Committee on Standards in Public Life (CSPL)
Dame Shirley Pearce	Independent Member, Committee on Standards in Public Life (CSPL)
Monisha Shah	Independent Member, Committee on Standards in Public Life (CSPL)
Dr Reuben Binns	Postdoctoral Fellow in AI, ICO
Dr Jonathan Bright	Senior Research Fellow and Political Scientist, Oxford Internet Institute, University of Oxford
Professor Lizzie Coles-Kemp	Professor for Information Security, Royal Holloway University
Emily Commander	Head of Public Policy, AHRC, UKRI
David Evans	Director of Public Affairs, Goodfaith
Professor Anthony Finkelstein	UK Government Chief Scientific Adviser for National Security, and Chair of Software Science, UCL
Professor Andrew Howes	Head of Computer Science, University of Birmingham
Professor Helen Margetts	Professor Helen Margetts, Director for Public Policy, The Alan Turing Institute
Dr Brent Mittelstadt	Research Fellow, Oxford Internet Institute, University of Oxford
Professor Paul Nightingale	Director of Strategy and Operations, Economic and Social Research Council
Dr John Sorabji	Principal Legal Adviser to the Lord Chief Justice and Master of the Rolls
Andrew Yell	Global Supplier Manager, Farnell



### Written submissions

The Committee received written submissions and additional written material from 20 individuals and organisations. No formal public consultation was held for this review.

Crofton Black, The Bureau of Investigative Journalism
Dr Emma Carmel, University of Bath
British Computer Society (BCS)
Carnegie UK Trust
Centre for Data Ethics and Innovation (CDEI)
Centre for Information Policy Leadership (CIPL)
Chartered Institute of Public Relations (CIPR)
Mission and Public Affairs Council, The Church of England
Robin Allen QC and Dee Masters, Cloisters Chambers
Ditto AI
David Evans, Good Faith Partnership
The Information Commissioner's Office (ICO)
Jamie Grace, Senior Lecturer in Law, Sheffield Hallam University
Dr Rune Nyrop, Dr Jess Whittlestone and Professor Stephen Cave, Leverhulme Centre for the Future of Intelligence, University of Cambridge
Christopher Marsh
MedConfidential
Marion Oswald, Senior Fellow, Department of Law, University of Winchester
The Royal College of Physicians
SAS Institute
Professor Karen Yeung, University of Birmingham

### Polling and focus group research

The Committee commissioned Deltapoll to run quantitative and qualitative research. Polling and focus groups examined attitudes towards AI. One focus group was held with members of the general public and one with front-line public sector officials. A report, focus group transcripts, and full data tables are available on the Committee's website.

#### DeltaPoll Survey Results

Sample Size: 2,016 GB Adults  
Fieldwork: 14-17 June 2019

**QV1** I would like you to think about the form of artificial intelligence (AI) that is advanced computer data analytics. This type of AI is computer software that analyses millions of data points of information, finds patterns within that data, and uses those patterns to come to a conclusion or insight about something within our world.

How comfortable or uncomfortable would you be, if at all, if decisions in the government and public sector were made using AI in each of the following scenarios?

**QV1\_1** AI is used to devise a care plan for a 7 year old with special educational needs.

Reaction	%
Very comfortable	9
Quite comfortable	25
Quite uncomfortable	30
Very uncomfortable	21
Don't know	15
Comfortable (All)	34
Uncomfortable (All)	51
Net comfort	-17

**QV1\_2** AI is used to evaluate if a prisoner should be released from jail, by predicting the chance the prisoner will reoffend.

Reaction	%
Very comfortable	6
Quite comfortable	16
Quite uncomfortable	28
Very uncomfortable	39
Don't Know	12
Comfortable (All)	22
Uncomfortable (All)	67
Net comfort	-45

**QV1\_3** AI is used to understand medical scans and diagnose cancer.

Reaction	%
Very comfortable	15
Quite comfortable	38
Quite uncomfortable	21
Very uncomfortable	15
Don't Know	11
Comfortable (All)	53
Uncomfortable (All)	36
Net comfort	+17

**QV1\_4** AI is used to identify fraud in immigration checks.

Reaction	%
Very comfortable	21
Quite comfortable	43
Quite uncomfortable	15
Very uncomfortable	11
Don't know	10
Comfortable (All)	64
Uncomfortable (All)	26
Net comfort	+38

**QV1\_5** AI is used to predict if petty criminals are likely to commit serious gun or knife crime.

Reaction	%
Very comfortable	12
Quite comfortable	28
Quite uncomfortable	26
Very uncomfortable	21
Don't know	12
Comfortable (All)	40
Uncomfortable (All)	47
Net comfort	-7



**QV1\_6** AI is used to scan CVs for unqualified applicants.

Reaction	%
Very comfortable	14
Quite comfortable	39
Quite uncomfortable	21
Very uncomfortable	12
Don't know	13
Comfortable (All)	53
Uncomfortable (All)	33
Net comfort	+20

**QV2** Thinking of the previous scenarios shown, how confident are you that the government and public sector will use AI in an ethical way?

Reaction	%
Very confident	5
Quite confident	26
Not very confident	37
Not confident at all	16
Don't know	15
Confident (All)	31
Not Confident (All)	53
Net Confidence	-22

**QV3** Thinking of the above scenarios, which, if any, of the following would make you more comfortable with AI being used?

**QV3\_1** There is an easy-to-understand explanation for the AI software's decision.

Reaction	%
This would make me much more comfortable with AI being used	14
This would make me a bit more comfortable with AI being used	37
This would make no difference to me	33
This would make me less comfortable with using AI	3
Don't know	13
More comfortable (All)	51

**QV3\_2** A human operator always has the final say on whether to accept or reject an AI decision.

Reaction	%
This would make me much more comfortable with AI being used	31
This would make me a bit more comfortable with AI being used	38
This would make no difference to me	18
This would make me less comfortable with using AI	4
Don't know	9
More comfortable (All)	69



**QV3\_3** The AI has been deemed acceptable by a government ethics regulator.

Reaction	%
This would make me much more comfortable with AI being used	10
This would make me a bit more comfortable with AI being used	26
This would make no difference to me	42
This would make me less comfortable with using AI	11
Don't know	11
More comfortable (All)	36

**QV3\_4** You have the right to appeal against an AI decision to a human specialist.

Reaction	%
This would make me much more comfortable with AI being used	25
This would make me a bit more comfortable with AI being used	41
This would make no difference to me	20
This would make me less comfortable with using AI	5
Don't know	9
More comfortable (All)	66

**QV3\_5** The AI is known to have a 95% accuracy and success rate.

Reaction	%
This would make me much more comfortable with AI being used	16
This would make me a bit more comfortable with AI being used	38
This would make no difference to me	28
This would make me less comfortable with using AI	7
Don't know	11
More comfortable (All)	54

**QV3\_6** You understand clearly how the AI works.

Reaction	%
This would make me much more comfortable with AI being used	19
This would make me a bit more comfortable with AI being used	32
This would make no difference to me	30
This would make me less comfortable with using AI	4
Don't know	14
More comfortable (All)	51

**Committee on Standards in Public Life**

Room GC.07, 1 Horse Guards Road, London, SW1A 2HQ

Tel: 020 7271 2685

Email: [public@public-standards.gov.uk](mailto:public@public-standards.gov.uk)

February 2020