REFORM

GETTING THE MACHINE LEARNING

Scaling AI in public services

Joe Hill Sean Eke September 2024

Reform is established as the leading Westminster think tank for public service reform. We believe that the State has a fundamental role to play in enabling individuals, families and communities to thrive. But our vision is one in which the State delivers only the services that it is best placed to deliver, within sound public finances, and where both decision-making and delivery is devolved to the most appropriate level. We are committed to driving systemic change that will deliver better outcomes for all.

We are determinedly independent and strictly non-party in our approach. This is reflected in our cross-party Advisory Board and our events programme which seeks to convene likeminded reformers from across the political spectrum.

Reform is a registered charity, the Reform Research Trust, charity no. 1103739.

ABOUT REIMAGINING THE STATE

After a decade of disruption, the country faces a moment of national reflection. For too long, Britain has been papering over the cracks in an outdated social and economic model, but while this may bring temporary respite, it doesn't fix the foundations. In 1942 Beveridge stated: "a revolutionary moment in the world's history is a time for revolutions, not for patching." 80 years on, and in the wake of a devastating national crisis, that statement once again rings true. Now is the time to fix Britain's foundations.

Reform's new programme, *Reimagining the State*, will put forward a bold new vision for the role and shape of the State. One that can create the conditions for strong, confident communities, dynamic, innovative markets, and transformative, sustainable public services.

Reimagining Whitehall is one of the major work streams within this programme.

ABOUT REIMAGINING WHITEHALL

This paper is part of the *Reimagining Whitehall* work stream. To effectively reimagine the State, major change must occur in the behaviours, processes, and structures of central government. This paper examines how the Government should use Whitehall to rapidly scale up adoption of Artificial Intelligence throughout public services. It provides a comprehensive set of recommendations for radically overhauling how the State approaches AI-powered technology with greater central leadership, more flexible funding, and a less risk-averse approach to testing.

Reimagining Whitehall Steering group

Reform is grateful to the expert members of the *Reimagining Whitehall Steering Group* who provide invaluable insight and advise on the programme. Their involvement does not equal endorsement of every argument or recommendation put forward.

Dr Henry Kippin, Chief Executive, North of Tyne Combined Authority

Sir Geoff Mulgan CBE, Former Head of Policy, Prime Minister's Office; former Director of the No.10 Strategy Unit

Philip Rycroft CB, Former Permanent Secretary, Department for Exiting the European Union **Professor Jonathan Slater**, Former Permanent Secretary, Department for Education

Rachel Wolf, Founding Partner, Public First; Co-Author, 2019 Conservative Manifesto

Rt Hon Baroness Smith of Malvern (resigned on entering government July 2024), Minister for Skills, former Home Secretary

ACKNOWLEDGEMENTS

External reviewers

We would like to express our gratitude to Fiona Deans, Ruth Kelly and Professor Mark Thompson for their helpful comments on an earlier draft of this paper.

The arguments and any errors that remain are the authors' and the authors' alone.

Interviewees

We would like to thank all 37 interviewees for giving their time and candid insights to support this research paper.

The list of interviewees is as follows:

- Mike Bracken, Founding Partner, Public Digital, former Executive Director of the UK Government Digital Service and former Government Chief Data Officer
- Dave Bradley, Head of Policy, Teach First
- Seb Barker, Co-Founder & Chief Operating Officer, Beam
- Jonathan Bright, Fellow and Head of AI for Public Services, Alan Turing Institute
- Alex Chalmers, Platform Lead, Air Street Capital
- Rupert Chaplin, Head of Data Science, NHS Digital
- Rachel Coldicutt, Executive Director, Careful Trouble
- Alan Davies, Executive Advisor, Health Innovation Kent Surrey Sussex
- Fiona Deans, Consultant and executive coach, former Chief Operating Officer, Government Digital Service.
- Joshua Entsminger, PhD student in innovation and public policy, Institute for Innovation and Public Purpose
- Richard Evans, Public Policy & Public Affairs Lead, Palantir
- Deb Fish, Research Fellow, Defence, Science and Technology Laboratory
- Tom Forth, Co-Founder and CTO, The Data City and Head of Data, Open Innovations
- Laura Gilbert, Director of Data Science at 10 Downing Street, & Director of the Incubator for AI, Cabinet Office.
- Gina Gill, Chief Strategy Officer, Central Digital and Data Office, Cabinet Office
- Andy Gregory, Deputy Director, Data Services and Analytics, Home Office
- Hugh Harvey, Managing Director, Hardian Health
- Scott Hayden, Head of Teaching, Learning and Digital, Basingstoke College of Technology
- Ben Henshall, Deputy Director Data Science & AI, 10 Downing Street
- Ruth Kelly, Chief Analyst, National Audit Office
- Simon King, Director of AI and Innovation, Department for Work and Pensions
- Paul Maltby, Director of Public Services, Faculty
- Archit Mehra, Data & Al Consulting Lead, Methods Analytics
- Kameswarie Nunna, Director Data Science, Simulation and AI, Newton Europe
- Sian Rodway, Chief Operating Officer, MDRx

- Richard Sargeant, Managing Director and Partner, Boston Consulting Group
- Renate Samson, Special Projects Lead, Ada Lovelace Institute
- Ryan Shea, Managing Director, PUBLIC
- Tom Shinner, COO, Entrepreneur First and former Director for Policy and Delivery Coordination at the Department for Exiting the European Union
- Simon Staffell, Director of Government Affairs, Microsoft
- Nora Stern, Co-Founder, Affiniti Al
- Mark Thompson, Professor in Digital Economy, University of Exeter

and six other interviewees who wished to remain anonymous.

METHODOLOGY

In addition to semi-structured interviews and desk research, this paper draws on the findings from responses to six Freedom of Information (FOI) requests, and procurement data aggregated by Tussell.

FOI requests

Four FOI requests were sent to each of 17 departments and one executive non-departmental public body listed below. These were:

- Cabinet Office
- Department for Business and Trade
- Department for Culture, Media and Sport
- Department for Education
- Department for Energy Security and Net Zero
- Department for Environment, Food & Rural Affairs
- Department for Levelling Up, Housing and Communities
- Department for Science, Innovation and Technology
- Department for Transport
- Department for Work and Pensions
- Department of Health and Social Care
- Foreign, Commonwealth & Development Office
- HM Revenue & Customs
- HM Treasury
- Home Office
- Ministry of Defence
- Ministry of Justice
- NHS England

A fifth FOI request was sent to each of the Department for Business and Trade, Department for Environment, Food & Rural Affairs, Department for Work and Pensions, HM Revenue & Customs, Home Office and the Ministry of Justice.

A sixth FOI request was sent only to the Cabinet Office.

Information on the requests and responses is contained in Appendix 3, published separately on *Reform's* website (reform.uk). The analysis is based on responses received up until 1 August 2024.

Procurement data

Tussell aggregates data on government contracts and spending. They provided *Reform* with breakdowns of this data for several categories of government procurement spending. The full set of analysis can be found in Appendix 2, published separately on *Reform's* website. More information about Tussell's data platform can be found at www.tussell.com.

Table of contents

1. INTRODUCTION	11
1.1 The potential of Al	11
1.2 Al in government	
1.3 The limitations of digital transformation in government	13
1.4 Getting the basics right and prioritising productivity	14
2. LEADERSHIP, DELIVERY AND GOVERNANCE	17
2.1 The digital centre of government	17
2.2 Leadership in departments, agencies and public bodies	18
2.3 Rebuilding central leadership	19
2.3.1 Leadership	19
2.3.2 Prioritisation	21
2.3.3 Oversight and control	
2.3.4 Guidance and support	23
3. FUNDING	
3.1 Realistically assessing the prize	25
3.2 State of play	37
3.3 Funding to scale	
3.4 Picking winners	32
4. BUILD IT, BUY IT, RENT IT	35
4.1 Who should develop AI for government?	35
4.2 Building it: in house capabiltiy	35
4.3 Buy it: better partnerships with the private sector	37
4.4 Rent it: using off-the-shelf and open-source products	40
5. DEVELOPMENT AND DEPLOYMENT	43
5.1 A principle-based approach	43
5.2 Data quality and access	
5.3 Risk assessment	49
5.3.1 Legal requirements	50
5.3.2 Data sharing	

5.4 Evaluating Al	54
5.4 Trust	55
6. CONCLUSION	58
APPENDIX 1: DEFINING AI	59
BIBLIOGRAPHY	60

Recommendations

Recommendation 1: The Government should establish a Government Data and Al Service (GDAIS), as a separate function within CDDO, sitting alongside the Government Digital Service. Its remit should be driving Al adoption across the public sector, and it should be led by a Government Chief Al Officer. The GDAIS should incorporate the current Incubator for Al (i.Al).

Recommendation 2: GDAIS should recruit Programme Directors to lead AI adoption, with one leading each of the priority use cases identified in this paper. The Programme Directors should work across all departments for which the use case is relevant, working in collaboration with the sector- or service-specific expertise held in that department.

Recommendation 3: The GDAIS should take on the responsibility for applying digital spend controls to all new spend with a contract value above £100,000, including spending on AI consultancy.

Recommendation 4: The Government Data & Al Service should be tasked with publishing guidance on Al adoption.

Within GDAIS, there should be:

- A team of specialists who serve as the point of contact with each department, understand its business and pipeline of projects, and advise on the spend control process.
- A central team which collects evaluations of AI from across government, and assures commonly used off-the-shelf and open source AI products on behalf of central government, making these resources available to government departments and public bodies.

Recommendation 5: Government should announce a new AI Transformation Fund, and HM Treasury should allocate GDAIS an additional budget of £1 billion between 2025-26 and 2027-28 to finance it.

The Fund should be available for projects which already have an existing evidence base, and demonstrably high productivity-boosting potential. There should be flexibility in what kinds of spending the Fund is used for, provided they support the scaling up of AI adoption in defined use cases.

Recommendation 6: HMT should agree an overall business case for investment with GDAIS, based on the Programme Directors' aggregate understanding of the value for money case for investment in their use cases. This business case should meet the recommendations of the Willets Review of the DSIT Business Case Process.

Recommendation 7: GDAIS Programme Directors should have the flexibility to allocate funding to projects in departments and public bodies which demonstrate the potential to represent good value for money, without reverting to HMT sign off. They should identify these opportunities by scouting them from those organisations.

Recommendation 8: Within funding they commit to projects, GDAIS Programme Directors should then have the flexibility to allocate funding to whatever capabilities are needed to deliver the project. This could include additional staff in GDAIS or the department, contracted support, software licenses, data curation and remediating legacy IT.

Recommendation 9: GDAIS should establish a cross-government Specialist Development Scheme for key AI roles, as outlined in the *Reform* paper *Making the grade*. To attract candidates to join the civil service on these terms, GDAIS should pre-agree pay frameworks for crucial AI roles with the Cabinet Office, which are at least in line with the levels paid for the AI Safety Institute and Incubator for AI.

Recommendation 10: GDAIS should create a new cross-government procurement framework for AI adoption in government, with different lots for supporting services. The framework should take advantage of the new Competitive Flexible Procedure to allow public bodies to bring in providers on a trial basis to test their services.

Recommendation 11: Each GDAIS Programme Director should establish an 'AI Sandbox' for the use case they lead, an environment to bring together public sector organisations and prospective partners to discuss ideas, meet prospective users and test sample data.

Recommendation 12: GDAIS should develop and run a new procurement framework exclusively designed for the procurement of off-the-shelf AI tools and enabling infrastructure, with an extremely light-touch initial process for trialling and experimenting with new products before going out to a full competitive process. This should be in recognition that the financial commitments required to experiment with products available via license fee are inherently much lower than other kinds of procurement, so less risky.

Recommendation 13: GDAIS's central guidance and support function should assess offthe-shelf and open source AI products and enabling infrastructure, and provide a list of approved ('kitemarked') products with supporting assessments to public bodies to support their procurement decisions.

Recommendation 14: GDAIS should produce and publish a single, simple set of principles for using AI in public services.

Recommendation 15: Departments and public bodies should publish their own policies, applying these central principles to their own specific circumstances.

Recommendation 16: Government legal advice, internal processes and documents (including Data Protection Impact Assessments) should give equal parity to the risks of not using AI or automated processing, as they do to the risks of using AI to partially or completely automate the process. This should be based on the performance and risks of the current system and processes used.

Recommendation 17: Government internal processes and documents for data sharing, including Data Sharing Agreements and DPIAs should require officials to also assess the risks of not sharing data, and give these equal parity with the risks involved in any new sharing.

Recommendation 18: The GDAIS should publish central government guidance on the principles of evaluating AI, including benchmarking the performance of AI against the performance of humans doing the same task, and evaluating against human performance on an ongoing basis once implemented via control groups.

Recommendation 19: Every central government body should appoint a single named Algorithm Owner, with appropriate AI literacy, for each AI algorithm in use in their organisation and provide them with core training on the requirements of their role.

1. Introduction

1.1 The potential of AI

Artificial Intelligence (AI) is fundamentally transforming the economy and society.

While machine learning models can make mistakes in classifying and predicting the future, and Generative AI models often 'hallucinate' and fail in other complex ways, AI is nevertheless surpassing human performance on many tasks.¹ AI capabilities continue to grow apace, underpinned by the huge growth in investment – more than \$25 billion was invested in the sector in 2023, nearly nine times the amount invested in 2022 and about 30 times that invested in 2019.²

The potential of AI to transform government is no less great. Much has been written about its potential applications and corresponding benefits, but at a time of performance and fiscal crisis, it is in boosting productivity that AI's greatest potential lies. Along with enabling public services to be more efficient, it can also relieve pressure on burdened services, automate low-value administrative tasks, and help government support more citizens, patients and users.

Public debt stands at 101.3 per cent of GDP,³ and the tax burden at a post-war high.⁴ But most worrying, despite spending increases, public services remain in a dire state. Pre-existing stagnant performance has been worsened by the effects of the Covid-19 pandemic, creating large backlogs and bottlenecks across the public sector. The Office for National Statistics (ONS) assesses public service productivity as still lower in 2023 than it was in 1997.⁵

The productivity challenge is complicated by Baumol's cost disease, which stipulates that labour-intensive services experience continually rising costs even while productivity remains stagnant.⁶ Applied to the public sector, this means that the cost of services will continue to rise to maintain the same level of output, and performance can only increase if spending does. Public services need a radical productivity boost which decouples them from this productivity trap.

Al is well placed to enable this transformative reset. The National Audit Office (NAO) reported an internal government assessment that a third of tasks in the civil service could be automated,⁷ and some estimates suggest that Al could provide £200 billion worth of public sector productivity improvements over the next five years.⁸ Previous work by *Reform* has

¹ Stanford Institute for Human-Centered AI, Artificial Intelligence Index Report 2024, 2024.

² Ibid.

³ Office for National Statistics, 'UK Government Debt and Deficit: December 2023', Web Page, 30 April 2024.

⁴ Office for Budget Responsibility, *Economic and Fiscal Outlook March 2024*, 2024.

⁵ Office for National Statistics, 'Public Service Productivity, Quarterly, UK: January to March 2024', Web Page, 15 July 2024.

⁶ William J. Baumol and William G. Bowen, *Performing Arts: The Economic Dilemma* (New York: The Twentieth Century Fund, 1966).

⁷ National Audit Office, Use of Artificial Intelligence in Government, 2024.

⁸ Alexander Iosad, David Railton, and Tom Westgarth, *Governing in the Age of AI: A New Model to Transform the State* (Tony Blair Institute for Global Change, 2024).

highlighted, for example, the scale of the opportunity for AI in public services such as the NHS.⁹ The prize, if executed well, is huge.

Previous debates about AI's role have been complicated by AI having no universally accepted definition, and the term is often used to refer to many different processes defined in software. For the purposes of this paper, *Reform* use the definition outlined by the Defence Science and Technology Laboratory (Dstl):

Theories and techniques developed to allow computer systems to perform tasks normally requiring human or biological intelligence.¹⁰

This is necessarily a broad definition of AI. It is broader than the well-publicised Generative AI models such as Chat-GPT. It also includes technologies which are common in most commercially-available software, including rules-based decision tools. Conflating these technologies is unavoidable, given the distinction between AI and other kinds of software will become more blurred over time.

More detailed definitions of common types of AI, and relevant categories like machine learning and data science, are provided in Appendix 1.

1.2 Al in government

The 2019-2024 Conservative Government recognised the opportunity presented by AI. New guidance and strategy documents have been issued, for example the *National AI Strategy* in 2021,¹¹ *A pro-innovation approach to AI regulation* in 2023,¹² and the *Generative AI Framework for HM Government* in 2024.¹³ And new bodies were established to attract data science expertise into government, for example the Incubator for Artificial Intelligence (i.AI) and the AI Safety Institute (AISI).

Overall, however, the State's focus has been on the contribution AI can make to economic growth and positioning the UK to have key capabilities to evaluate AI safety.

These are undeniably important. Adopting AI throughout the private sector is crucial to unlocking economic growth. Research suggests that wide integration of AI across the economy could deliver an economic boost of 13 per cent of GDP,¹⁴ and a five year delay in adoption could reduce the size of the economic impact by more than £150 billion by 2035.¹⁵ Likewise, UK-based capacity to evaluate the safety of frontier AI models will be similarly crucial to future economic success, geopolitical security and social cohesion.

However, the use of AI to drive public sector productivity, despite considerable interest, remains a somewhat neglected area. Early measures to establish the Incubator for AI (i.AI),

⁹ Eleonora Harwich and Kate Laylock, *Thinking on Its Own: AI in the NHS* (Reform, 2018).

¹⁰ Defence, Science and Technology Laboratory, *The Dstl Biscuit Book: Artificial Intelligence, Data Science and (Mostly) Machine Learning*, 2019.

¹¹ HM Government, *National AI Strategy*, 2021.

¹² Department for Science, Innovation and Technology, *A Pro-Innovation Approach to AI Regulation*, 2023.

¹³ HM Government and Central Digital and Data Office, *Generative AI Framework for HM Government*, 2024.

¹⁴ Carsten Jung and Bhargav Srinivasa Desikan, *Transformed by AI* (Institute for Public Policy Research, 2024).

¹⁵ Microsoft and Public First, *Unlocking the UK's AI Potential*, 2024.

publish the Generative AI Framework, and widen the remit of CDDO and the Responsible Technology Adoption Unit (now part of the AI Opportunities Unit) are promising.¹⁶ But i.AI is too small to drive significant 'top-down' technological change, the guidance disincentivises 'bottom-up' AI adoption, and wider remits for CDDO and the AI Opportunities Unit will only work if they are appropriately resourced.

1.3 The limitations of digital transformation in government

The Government has a mixed track record of technological adoption. Successes include the establishment of GOV.UK by the Government Digital Service (GDS) in 2012, which unified hundreds of different public-facing applications into shared platforms. The Driver and Vehicle Licensing Agency (DVLA) was one of the first government bodies to bring its IT in-house, and during the Covid-19 pandemic they managed to digitise the service to change addresses on vehicle log books. The new process took six weeks compared to the previous six months.¹⁷

Furthermore, the pandemic highlighted the ability for fast technological innovation at scale, for example the Covid-19 Early Warning System fed by data from NHS trusts across the country; the mass testing programme which at its peak was processing more than 100,000 tests a day;¹⁸ and the NHS App, which is estimated to have prevented around one million cases, 44,000 hospitalisations and 9,600 deaths in its first year.¹⁹

Despite these successes, failures are much more common. Examples include the NHS National Programme for IT, which is estimated to have cost upwards of £9.8 billion;²⁰ the introduction of the National Law Enforcement Data Service (NLEDS), which has seen its costs increase by 68 per cent and is long overdue to replace the Police National Computer;²¹ and the Digital Services at the Border (DSAB) programme, which has been delayed by three years and total costs have increased by £173 million.²² These challenges are well known, yet as the NAO characterise it, "despite 25 years of government strategies and countless attempts to deliver digital business change successfully, our reports show a consistent pattern of underperformance".²³

That is because the root causes of these and other failures remain. Legacy IT systems make technological change harder to achieve without disruption,²⁴ and data contained within government systems is often of poor quality and hard to access.²⁵ While some progress has been made on skills, with over 50 per cent of Fast Stream hires having STEM degrees in

¹⁶ Department for Science, Innovation and Technology, 'AI Expert to Lead Action Plan to Ensure UK Reaps the Benefits of Artificial Intelligence', Web Page, 26 July 2024.

¹⁷ Driver and Vehicle Licensing Agency, 'DLVA's IT Transformation – Our Story so Far', Web Page, 3 August 2022.

¹⁸ Department of Health and Social Care, 'Daily Tests Processed and Testing Capacity (UK): 20 March to 22 September 2020', Web Page, 24 September 2020.

¹⁹ Michelle Kendall et al., 'Epidemiological Impacts of the NHS COVID-19 App in England and Wales Throughout Its First Year', *Nature Communications* 14 (2023).

²⁰ Public Accounts Committee, *The Dismantled National Programme for IT in the NHS*, 2013.

²¹ National Audit Office, 'The National Law Enforcement Data Programme', Web Page, 10 September 2021.

²² Public Accounts Committee, *Digital Services at the Border*, 2021.

²³ National Audit Office, *The Challenges in Implementing Digital Change*, 2021.

²⁴ Financial Conduct Authority, *Implementing Technology Change*, 2021.

²⁵ National Audit Office, Challenges in Using Data Across Government, 2019.

2023,²⁶ and government's Digital and Data Profession grew by 19 per cent between April 2022 and April 2023,²⁷ there remains a widely recognised shortage of digital and technological expertise – less than 4 per cent of civil servants are digital professionals compared with an industry average of between 8 and 12 per cent.²⁸ And many of the barriers to reform that impede Whitehall delivery more generally also apply to digital programmes, including a lack of clarity over who is responsible for change and insufficient investment in change management.²⁹

1.4 Getting the basics right and prioritising productivity

Al can fundamentally transform public services, many of which could be rebuilt from the ground-up to take advantage of the emerging technologies in this field. However, in many areas the State is not ready to make that kind of radical transformation, and an iterative approach is essential to see short-term benefits to services which are struggling to meet current demands.

Recent proposals to government — for example from Labour for the Long Term³⁰ and the Tony Blair Institute³¹ — have focused on investment in sovereign foundational AI models (e.g. Large Language Models) to keep pace with a rapidly advancing technology, and mitigate the risks of losing access to models provided by companies based in the United States. The risks of this happening are hard to assess, as is the Government's ability to build models which could reasonably compete with rapidly-improving foundational models built by the private sector.

Given government's mixed record on technological adoption, and the scale of the fiscal and performance challenges in public services, it would be prudent for Whitehall to focus on scaling up the deployment of AI applications which are already well tested but not yet integrated in the public sector. In short, rather than experimenting with fast evolving, cutting-edge AI, the core focus in government should be on applying and scaling those technologies that can have the biggest, near-term impact on productivity.

Specifically that means government should ruthlessly prioritise deploying AI in areas that:

- (a) have an existing evidence base for effectiveness; and
- (b) can realise productivity benefits within two or three years of deployment.

The use cases listed below in Figure 1 fit these criteria.

²⁶ Cabinet Office, 'Civil Service Fast Stream Exceeds Target to Boost STEM Expertise Across Government', Web Page, 25 October 2023.

²⁷ Central Digital and Data Office, *Transforming for a Digital Future:* 2022 to 2025 Roadmap for Digital and Data - Updated September 2023, 2023.

²⁸ National Audit Office, *Digital Transformation in Government: Addressing the Barriers to Efficiency*, 2023.

²⁹ Charlotte Pickles and James Sweetland, *Breaking Down the Barriers: Why Whitehall Is so Hard to Reform* (Reform, 2023).

³⁰ Haydn Belfield, *Great British Cloud and BritGPT: The UK's AI Industrial Strategy Must Play to Our Ambitions* (Labour for the Long Term, 2023).

³¹ losad, Railton, and Westgarth, Governing in the Age of Al: A New Model to Transform the State.

|--|

Use case	Description	Size of the prize	State of play	Examples
Business planning	Mapping and predicting demand for public services at a business level and planning how to resource it. Examples include highlighting crime hotspot areas/times and managing A&E priority lists.	Research suggests that effective targeting of crime hotspots can reduce violence against the person and robbery offences by 7 per cent. ³³	There is signification variation in how this business planning is done in different public sector bodies and parts of public services. For example, several police forces have experimented with using predictive algorithms ³⁴ and some NHS trusts have experimented with algorithms to support triaging patients. ³⁵ It is unlikely that one standardised approach would work across all sectors.	From October 2022 to September 2023 AI was applied to the footage from every camera in Willesden Green Tube Station. The AI had the ability to identify up to 77 different use cases, ranging from significant incidents, for example fare evasion, to more minor occurrences, for example litter. 19,000 alerts were issued in real time to station staff to act on. ³⁶ PredPol is a company which uses algorithms to predict crime locations. During a four month trial using the software, Kent Police achieved a 6 per cent reduction in crime, and PredPol claims that it can improve crime detection by 10 to 50 per cent in some cities. ³⁷
Assessment streamlining	Increasing the speed at which decisions can be made. Examples include processing asylum claims, Universal Credit Claims, and assessing the outputs of diagnostic tests (e.g. examining chest X-rays).	In December 2022 there were 1,237 caseworkers working on asylum claims. On average, every member of staff made only four decisions a month. ³⁸ In April 2023 there were 1.6 million people in England waiting for a diagnostic test and 431,000 had been waiting more than six weeks. ³⁹	Decisions in government are currently mainly made by specialist professionals, consulting data and guidance. Some AI applications which support them are being trialled, for example the AI Diagnostic Fund is providing £21 million in funding to speed up the roll out of AI diagnostic tools across 64 NHS trusts. ⁴⁰ And in May 2023 a 'hackathon' was organised by the Home Office to explore possible ways AI could help reduce the asylum backlog. ⁴¹	In Jobcentres, an AI tool called 'a-cubed' trawls thousands of pieces of guidance to give work coaches information on the best support to help claimants into work quicker. ⁴² AQA are trialling whether AI can be used to provide "quality assurance" to humans marking GCSE and A-level exams. ⁴³

³² Other use cases *Reform* researched demonstrated some evidence base, such as the use of AI tools to support document retrieval and analysis in a policy context, however the benefits case for widespread automation of these functions is not yet well established.

³³ Olivia Jeffrey et al., Evaluation Report on Grip and Bespoke-Funded Hot Spot Policing (Home Office, 2024).

³⁴ Patricia Nilsson, 'First UK Police Force to Try Predictive Policing Ends Contract', *Financial Times*, 26 November 2018.

³⁵ The NHS AI Lab, 'Using AI to Improve Back Office Efficiency in the NHS', Web Page, 11 February 2022.

³⁶ Matt Burgess, 'London Underground Is Testing Real-Time AI Surveillance Tools to Spot Crime', *WIRED*, 8 February 2024.

³⁷ Mark Smith, 'Can We Predict When and Where a Crime Will Take Place?', *BBC News*, 30 October 2018.

³⁸ Tom Sasse, Rhys Clyne, and Sachin Savur, 'Asylum Backlog', Web Page, 24 February 2023.

³⁹ British Medical Association, 'NHS Diagnostics Data Analysis', Web Page, 4 October 2023.

⁴⁰ The NHS AI Lab, 'AI Diagnostic Fund', Web Page

⁴¹ Amina Memon et al., 'Artificial Intelligence (AI) in the Asylum System', *Medical, Science and the Law* 64, no. 2 (2024).

⁴² Isabella McRae, 'DWP Ramps up AI Use to "Bring the Future to the Welfare System" and Push People Into Work', *The Big Issue*, 21 May 2024.

⁴³ Cerys Turner, 'GCSEs 2024: Exam Board to Trial AI in Summer Exams', *TES Magazine*, 15 March 2024.

GETTING THE MACHINE LEARNING

			Evaluation will be needed for effectiveness and bias in different case working environments.	
Process automation	Automating basic tasks. Examples include checking documents uploaded to GOV.UK and exam marking.	Exam markers for OCR are paid between £240 and £1,500 for a full allocation of papers (between 200 and 450 scripts). ⁴⁴	No consistent use of AI, with two major barriers being legacy infrastructure and the funding to replace it.	NHS England's Transformation Directorate has published guidance on applying robotic process automation in health. ⁴⁵ HM Land Registry partially automates the process for assessing applications to change the registry, and over 40 per cent of applications are completed within minutes. ⁴⁶
Chatbots & automated call centres	Providing AI chatbots and call centres.	More than 35,000 people work for the Department for Work and Pensions' in-house contact centre.	A growing number of web chats and call centres are handled or partially handled by AI in the private sector, but public sector adoption is slower.	In 2023 Derby City Council deployed phone-based AI and chatbots Darcie and Ali. Derby has handled 43 per cent of traditional phone conversations automatically, exceeding the target of 20 per cent. ⁴⁷
Translation/ transcription	Providing AI translation and transcription services. Examples include real-time translation of asylum, social care and criminal justice interviews.	The NHS spent at least £114 million on interpreters and translation services from 2019-20 to 2021-22. ⁴⁸ London Boroughs rated transcription of meeting notes as the highest priority use case for AI in social care. ⁴⁹	Multiple AI translation and transcription companies operate outside of government, however there is no evidence of uptake in central government. It will be important to prove translation services are effective for sensitive and technical details (e.g. medical symptoms).	Haringey Council has deployed AI to translate documents, reducing the cost of a ten page translation from £120 to 21p. ⁵⁰ The gastroenterology team at King's College Hospital have used Hepian's AI-driven platform to provide a transcription service. ⁵¹ This has decreased the time doctors spend writing letters, allowing them to refocus on more valuable tasks.
Coding co- pilots	Assisting developers with writing code.	Digital, Data and Technology is the third largest government profession after Operational Delivery and Policy, with a headcount of 23,155 (4.7 per cent).	It is unclear how much internally developed government code is produced using co-pilots. The <i>Generative AI Framework for HM</i> <i>Government</i> advises civil servants to review the output of any co-pilot models to ensure it does not breach intellectual property law. ⁵²	GitHub Copilot provides assistance to government software developers writing code. It has been used by more than one million developers and generated over three billion lines of accepted code. Users accept nearly 30 per cent of code suggestions from GitHub, increasing productivity by reducing the time taken writing bespoke code. ⁵³

⁴⁴ OCR, 'Examiners', Web Page.

- ⁴⁶ HM Land Registry, 'HM Land Registry: Processing Times', 19 July 2024.
- ⁴⁷ Microsoft, Harnessing the Power of AI for the Public Sector, 2024.
- ⁴⁸ TaxPayers' Alliance, NHS Spending on Interpreters and Translation Services, 2023.
- ⁴⁹ London Office of Technology and Innovation & Faculty, *Opportunities for AI in Adult Social Care Services*, 2024.
- ⁵⁰ Microsoft, *Harnessing the Power of AI for the Public Sector*.
- ⁵¹ NHS England, 'Using an AI-Driven Dictation Platform to Free up Clinicians' Time', Web Page, n.d.
- ⁵² HM Government and Central Digital and Data Office, *Generative AI Framework for HM Government*.
- ⁵³ Thomas Dohmke, 'The Economic Impact of the AI-Powered Developer Lifecycle and Lessons From GitHub CoPilot', Web Page, 14 May 2024.

⁴⁵ McRae, 'DWP Ramps up AI Use to "Bring the Future to the Welfare System" and Push People Into Work'.

2. Leadership, delivery and governance

There is no clear accountability for driving AI adoption across government, and responsibility for leading the agenda is split between multiple bodies in Whitehall. The implementation of digital change is almost entirely devolved to Whitehall departments, local authorities, NHS services, police forces and multi-academy trusts. Central government, now led by the Department for Science, Innovation and Technology (DSIT), currently plays a very limited role in setting standards, piloting projects and delivering some enterprise services across government. The new AI Opportunities Unit, bringing together previously separate teams in DSIT (including i.AI, which has moved from the Cabinet Office) could be an opportunity to lead more from the centre.

It is not possible to centralise all AI adoption across the public sector. But to radically scale up AI in government, more central leadership is needed to address the systemic barriers to AI adoption.

2.1 The digital centre of government

Central government leadership of AI in public services is underpowered to deliver change topdown across large parts of public services. Establishing i.AI is a good start, but with its current level of resourcing it is insufficient to drive overall cross-government adoption. With around 70 staff and a budget of £101 million over five years, its focus to-date has been on internal consultancy and supporting departments to pilot new projects.⁵⁴

Alongside this, the Central Digital and Data Office (CDDO) oversees government strategy, sets common digital and technology standards, and produces central guidance to support the public sector in AI adoption – for example, in publishing the government's Generative AI Framework in 2024. The Government Digital Service (GDS) provides common digital products such as GOV.UK to other government departments, including instances where AI may need to be integrated into those enterprise systems.⁵⁵

Under the new Labour Government, i.AI, CDDO and GDS have all been moved from the Cabinet Office into the Department for Science, Innovation and Technology (DSIT) – locating them alongside the policy functions responsible for AI policy within the wider economy, and R&D funding for private and public AI initiatives.⁵⁶

Interviewees for the paper, and participants in *Reform* research roundtables, were optimistic that this move would be positive for AI adoption in public services. Encouragingly, one interviewee said "DSIT Ministers seem to place a clear priority on digital in government".

This is important, and many case studies, as well as interviewees, highlight the role that former Cabinet Office minister Frances Maude played in the success of GDS in its early days.⁵⁷ One

⁵⁴ National Audit Office, Use of Artificial Intelligence in Government.

⁵⁵ Ibid.

⁵⁶ Department for Science, Innovation and Technology, 'DSIT Bolstered to Better Serve the British Public Through Science and Technology', Web Page, 8 July 2024.

⁵⁷ Paul Maltby, 'Reasons to Be Optimistic About DSIT and the New Digital Centre of Government', Web Page, 24 July 2024.

interviewee noted in particular the importance of sustained political leadership by a figure with deep private sector expertise: Maude was responsible for digitising public services for five years between 2010 and 2015, and so "had the clear mandate to force through difficult changes in how departments worked with each other". Another noted "how important Francis Maude was", lamenting that "since then, most ministers haven't known what GDS did".

However, rather than focusing on the case for locating AI leadership in DSIT, most interviewees focused on the disadvantages of retaining it in the Cabinet Office, a department which one described as "overstretched... trying to do forty different things, of which digital is just one". In previous *Reform* research, one senior interviewee characterised the Cabinet Office as a "uniquely dysfunctional organisation", while one former permanent secretary argued it isn't "an 'out there'... sort of organisation. In its central function, it is solving problems for the Cabinet Office, not for departments."⁵⁸

This calls into question its power as a central 'convening force' in government, and suggests an organisation ill-suited to the responsibility of leading AI adoption throughout public services. As one former GDS leader has written, "a Cabinet Office email address alone is insufficient to imprint effective and strong leadership over Whitehall."⁵⁹

The strongest case for moving AI responsibility to DSIT (along with digital transformation more broadly) is the concentration of scarce talent in one location. According to one civil servant interviewed for this paper, "CDDO and i.AI will have more colleagues who are actually interested and knowledgeable about technology", contrasting it with a Cabinet Office culture which was only interested in "whether you could reduce your team's headcount by five people next year". Government AI capabilities should remain in DSIT.

2.2 Leadership in departments, agencies and public bodies

Most operational decisions about how to build and deploy software (including AI) are made by public bodies themselves – government departments, local authorities, NHS organisations, schools and police forces, not to mention a long list of non-departmental public bodies (NDPBs). The absence of a substantial, capable centre driving AI deployment means nearly all the delivery risk lies with these bodies, "with ultimate responsibility for delivery residing with the Permanent Secretary or Accounting Officer in each department."⁶⁰ Last year, the government established an AI Directors' Policy Board, comprised of nominated Directors responsible for AI in each government department. Its remit includes policy development, risk management, and transparency and coordination.⁶¹

However, there is little evidence to support the belief that radical transformation through AI will happen organically. Given the State's bottom-up digital transformation has been slow – with only ten of government's 'top 75' services deemed to be at a 'great' standard and 45 requiring

⁶¹ See Appendix 3

⁵⁸ Pickles and Sweetland, *Breaking Down the Barriers: Why Whitehall Is so Hard to Reform*.

⁵⁹ Maltby, 'Reasons to Be Optimistic About DSIT and the New Digitial Centre of Government'.

⁶⁰ Central Digital and Data Office, *Transforming for a Digital Future:* 2022 to 2025 Roadmap for Digital and Data - Updated September 2023.

'significant improvement'⁶² – it is unlikely that AI adoption will be any swifter if it works the same way.

Whilst Generative AI models entered the public consciousness in 2022 with the early demo of ChatGPT, other kinds of machine learning software have been used in public services for many years. The Department for Work and Pensions (DWP) set up their Intelligent Automation Garage in 2017 to grow automation across the department.⁶³ Yet despite this awareness, as Marc Warner, CEO of Faculty, put it in a panel discussion, "the State is still basically nowhere on AI, it is still struggling with the more basic digital revolution".⁶⁴ Interviewees were particularly critical of the incentives for government departments and public bodies to transform themselves – as one interviewee put it:

If you're a leader in [department], you might be able to make your department more efficient and cheaper through AI. But you would have to take a lot of risk to do that, it might not pay off. Instead, why wouldn't you just ask for more money in next year's budget to improve frontline services?

Yet it is inevitable that most change will happen this way – the sheer breadth of public services means no central body can drive all change, and most public services have their own digital teams, albeit with varying capabilities. Encouragingly, multiple interviewees were keen to contrast how different the present day is to the early days of GDS, when "most of government didn't have in-house capability, or was locked into bad contracts to use private companies' technology".

Over time, more of the public sector will adopt AI, but without reform it is likely to be at too slow a pace. Many of the recommendations throughout this paper are thus aimed at overhauling the incentives for government bodies to adopt AI locally, in order to speed up the implementation, at scale, of tried and tested productivity boosting technology.

2.3 Rebuilding central leadership

All this means that a government which is determined to see more Al adoption will need more capability centrally.

Our research identified four roles the digital centre of government should play in driving Al adoption across public services:

- 1. Leadership
- 2. Prioritisation
- 3. Oversight and control
- 4. Guidance and support

2.3.1 Leadership

i.AI, CDDO and GDS together provide a connected but complex landscape of different organisations supporting Al adoption at the centre of government.

⁶² Public Accounts Committee, *Digital Transformation in Government: Addressing the Barriers to Efficiency*, 2023.

⁶³ DWP Digital, 'Intelligent Automation Garage: Using Automation for Good', Web Page.

⁶⁴ Civic Future, 'Should the State Play More of a Role in the Economy?', Web Page, 29 June 2024.

The GDS was originally set up in 2011 based on Martha Lane Fox's review of Directgov (the precursor to GOV.UK), which recommended the creation of what became GDS in response to "the reinvention of the internet and behaviour of users".⁶⁵ It aspired to a "service culture" in government, which focused on the needs of citizens rather than government departments.

GDS delivered substantial progress in a short space of time, with many examples in common use today. Its pioneering impact is evidenced by its international influence: the GDS model was the direct inspiration for the United States Digital Service,⁶⁶ and similar visions to its founding principles are echoed in the Singaporean Open Government Products team⁶⁷ and the Canadian Digital Service.⁶⁸ While some counties, such as Estonia and New Zealand, have had successful digital transformation without a central digital unit, these faced fewer obstacles with legacy IT infrastructure or found other ways to create the same principles of a central digital unit.⁶⁹

Separating GDS into two functions, by creating CDDO in parallel and splitting responsibility for delivery and for governance was intended to give clarity. It came from a 2020 report by the Digital Economy Council, which suggested that GDS had "to some extent lost its way" after going through several different incarnations since it was established in 2011.⁷⁰

In his 2023 review of civil service governance, Francis Maude said this caused confusion:

The centre is now frequently providing multiple (and mixed) signals. In digital, for example, the split between the Government Digital Service (GDS) and Central Digital and Data Office (CDDO) creates a largely artificial split between functional leadership and delivery. The lack of a unified organisational structure degrades the strength of leadership that can be provided by the centre, and absorbs significant amounts of officials' time in brokering internal coordination rather than delivery.⁷¹

Some interviewees felt the establishment of i.Al in 2023, whilst a very welcome capacity boost, complicated things further by providing a central leadership function for Al that is separate to these other areas of digital services. Some argued that Al shouldn't be seen as separate to overall digital service transformation: "it's just a different kind of software, it still needs to follow the same core principles of digital in government – being centred around the user, and developed as a service".

Another interviewee pointed out that AI faced exactly the same delivery challenges as earlier digital technologies continued to, and so should be left to the same teams to lead: "regardless of the technology, the issues don't change. In the 2010s, GDS was trying to 'fix the plumbing', get the right skills in, and set standards. AI is exactly the same".

Indeed, many thought separating AI from other digital service delivery came with risks – particularly the risk of "AI being the solution in search of a problem". "AI doesn't need different things from digital technology in government", argued one interviewee, "we always get excited

⁶⁵ Martha Lane Fox, *Directgov 2010 and Beyond: Revolution Not Evolution*, 2010.

⁶⁶ Tim O'Reilly, 'The UK Needs to Double Down on the GDS, Not Dismantle It', Web Page, 16 August 2016.

⁶⁷ Open Government Products, 'About Us', Web Page, n.d.

⁶⁸ Canadian Digital Service, 'Digital Products for the Government of Canada', Web Page, n.d.

⁶⁹ Andrew Bennett and Chris Yiu, *Transforming Government for the 21st Century* (Tony Blair Institute for Global Change, 2019).

⁷⁰ National Audit Office, *Digital Transformation in Government: Addressing the Barriers to Efficiency*.

⁷¹ Francis Maude, Independent Review of Governance and Accountability in the Civil Service, 2023.

by new technologies – in five years, it will be robotics". Another interviewee told said that "a lot of Generative AI is hype, and willingness to get that into government seems misguided", and that in contrast other kinds of intelligent software "like rules-based technology and neural networks, are already used in parts of government. Some parts have been using those for some time now, largely without incident".

The same interviewee told said that the kinds of AI which were in use in government were "mostly DevOps, most of the time. Once you manage data and build models well, and start to get the feedback from people, it stops becoming a magic black box and becomes about laborious processes of data management, testing and learning" – in other words, not an innovation but a routine part of digital transformation.

The overarching message from interviews is that AI development in government, and therefore i.AI, should not have separate leadership to wider digital development, and that the separation of CDDO and GDS confuses the accountability of leading the system and deploying services. However it is equally clear that AI should retain a distinct focus from the kind of cross-cutting digital services GDS now offers, because the kinds of AI used in government are not yet as mature as GOV.UK or One Login, and need to be built from the ground up.

This demands a different kind of leadership role. Government no longer needs an entrepreneurial GDS to drive overall digital transformation, because it has much more inhouse digital capability, but it does lack the specific capability to drive in-house AI adoption, and needs an approach like the early years of GDS. As one interviewee told us: "It is false to claim that we either need dedicated AI programmes in government, or to just leave AI to the existing digital programmes. The government's traditional digital approach needs to be updated to catch up with how AI has changed the world". And given the centrality of improving the quality and accessibility of government data (see Chapter 5), data governance policy should not be separated from leading AI adoption.

Within DSIT, teams transforming government digital services should be brought together to provide clear central leadership. Within DSIT's overall leadership of the system, GDS should continue its focus on maintaining shared digital products on behalf of the government. And to build on the initial work GDS and CDDO have done to support the AI adoption, a Government Data and AI Service (GDAIS) should be set up within CDDO but separate to GDS, to lead government's efforts to adopt AI across the public sector. This should be built out from the existing i.AI team, and GDAIS' internal leadership function should work with the AI Opportunities Unit to delivery of public sector commitments in the government's upcoming AI Opportunities Action Plan.⁷²

Recommendation 1: The Government should establish a Government Data and Al Service (GDAIS), as a separate function within CDDO, sitting alongside the Government Digital Service. Its remit should be driving Al adoption across the public sector, and it should be led by a Government Chief Al Officer. The GDAIS should incorporate the current Incubator for Al (i.Al).

⁷² Department for Science, Innovation and Technology, 'AI Expert to Lead Action Plan to Ensure UK Reaps the Benefits of Artificial Intelligence'.

2.3.2 Prioritisation

The leadership of AI in government should force ruthless prioritisation of the use cases outlined in Chapter 1, where, if successfully integrated, proven technologies can feasibly deliver sizeable impact in a short time frame. Trade-offs may need to be made between the scale of the potential in these areas, and how amenable they are to change.

Some use cases are entirely within the gift of Whitehall to operationalise, such as AI generated summaries of policy advice for Ministers, but represent relatively low overall economic value ("tinkering around the edges", as one interviewee put it). However, there are some areas where public services are controlled directly by central government, and represent big opportunities for productivity improvement via AI – the large operational functions of DWP, HM Revenue& Customs (HMRC) and the Home Office in particular. There are then other high-impact opportunities in local public services, where central government has fewer levers to drive transformation top-down.

A leadership structure for AI in government needs to be able to drive change in both of these, although the scale of opportunity and levers available in each case will determine the kinds of AI it can feasibly deliver.

A digital centre of government with clear priorities does not need to detract from the efforts of the rest of government. GDAIS can still let "1,000 AI pilots flourish" across the public sector,⁷³ while being rigorous in focusing its time and investment on those ideas which matter most to central government – pilots are only valuable if they can be scaled up into business-as-usual services.

This requires a model which is laser-focused on those opportunities which are identified as having the greatest evidence base and realisable benefits (the tests identified in Chapter 1). GDAIS should recruit Programme Directors who are focused on cross-cutting use, rather than on a sector basis (e.g. criminal justice, tax). This approach will provide the focus and drive required to identify and scale those technologies across government, rather than relying on each Programme Director to be a subject-matter expert in the many different kinds of use case which a given department or area of public services would have.

Recommendation 2: GDAIS should recruit Programme Directors to lead AI adoption, with one leading each of the priority use cases identified in this paper. The Programme Directors should work across all departments for which the use case is relevant, working in collaboration with the sector- or service-specific expertise held in that department.

2.3.3 Oversight and control

It is clear that, even within departments, there is little oversight and control of how AI is contributing to a future vision of public services. Here is a particularly stark example from one interviewee:

We were working with [government agency] to build a machine-learning triage tool, and their senior leadership team asked us to come in and talk about the future of AI and how it could change their business. On the same day we went

⁷³ Christopher McKeon, 'Let 1,000 AI Pilots Flourish to Help Public Sector, Says Labour MP', *Independent*, 9 August 2023.

in to talk to their board, [government agency] abruptly emailed us to tell us they were cancelling the project due to budget cuts. Their own leadership didn't know.

Several people interviewed mentioned the role that 'spend control' processes played in the early years of GDS, to ensure public spending wasn't being wasted on contracts which were not aligned with the vision of shared services. Spend control processes are still used, applied to spending above £100,000 on digital and £1,000,000 on technology which is not business as usual. These, along with any which meet criteria such as being "novel and contentious", go into a joint assurance process between the department and CDDO.⁷⁴ CDDO is already developing its spend controls to identify high-risk AI use cases, and ensure these are given appropriate scrutiny.⁷⁵

Interviewees stated that historically these controls were exercised much more strictly than they are today. Spend control processes should be continued and moved into the new GDAIS, so that it can ensure digital technology being developed is in line with the government's strategy for data and AI adoption – for example, common approaches to standards and interoperability, evaluation of model performance, and the right balance of procurement and in-house delivery.

Spend controls should not only be a tool for ensuring that spending is aligned with the government's technical objectives, but also that it is focused on those suppliers who are best suited to the kind of technical innovation required to implement AI. Two interviewees highlighted that AI development work was sometimes tendered via large consultancy frameworks like MCF3, which does not explicitly mention Artificial Intelligence in any of the nine lot descriptions.⁷⁶ The same interviewees felt this was contracted in this way because it was convenient for officials to use those providers, rather than the providers being best-placed to build AI software. These consultancy frameworks should also be in scope of GDAIS spend controls when they are used for technology consultancy.

Recommendation 3: The Government Data & AI Service should take on the responsibility for applying digital spend controls to all new spend with a contract value above £100,000, including spending on AI consultancy.

2.3.4 Guidance and support

Despite the diminished role of GDS, one of its most significant legacies, as one interviewee stressed, is the clear set of standards and guidance it produced, which diffused across central, and ultimately much of local, government. The GDS Service Manual⁷⁷ exists to support public sector digital teams in meeting the requirements of the Service Standard.⁷⁸

In contrast, while AI adoption requires different technical solutions to products like GOV.UK and One Login, government lacks the same clear guidance for teams adopting AI in public

⁷⁴ Central Digital and Data Office, 'Digital and Technology Spend Controls (Version 5)', Web Page, 21 July 2023.

⁷⁵ National Audit Office, Use of Artificial Intelligence in Government.

⁷⁶ Crown Commercial Service, *Management Consultancy Framework Three (MCF3)*, 2021.

⁷⁷ Government Digital Service, 'Service Manual', n.d.

⁷⁸ Government Digital Service, 'Service Standard', n.d.

services as the Service Standard which supports those products; a topic further discussed in Chapter 5. Such guidance is best produced centrally and adapted in departments within central government – and ideally the model is also powerful enough to influence local government.

Given the shortage of specialist AI, digital and technical skills in government, a central pool of resource to support government departments with AI adoption is best located in GDAIS. This could also be a forum for spreading best practice and learning. One interviewee said they were convinced that of the many pilot initiatives being discussed across government, plenty had already been tried, but because learning was not diffused throughout the system, those projects were likely to be repeated again by accident.

This central pool of resource would also provide capacity to challenge many of the structural barriers to AI adoption discussed in this paper – around technical development capacity, procurement, financing, performance and evaluation. Martha Lane Fox's original call for GDS involved "SWAT teams" who would be "given a remit to support and challenge departments and agencies delivering the first wave of digital only services".⁷⁹ Today this is needed for AI.

In order to professionalise the current teams and networks focused on AI adoption in government, GDAIS should become the professional home for AI in government. Interviewees argued that AI was still largely seen as the domain of science and analysts in government, given it is still largely funded from public R&D budgets at pilot stage. However, as projects scale into live services, different kinds of skills are required to manage them to those needed in the early stages of prototyping solutions, particularly when it comes to integrating AI into the large, operational services.

Given the strength of feeling among interviewees that AI should not be seen as separate to wider digital service transformation, it is clear that the Digital, Data and Technology profession, now led by CDDO, is the best home for professional support for AI experts within government, with input into professional standards and training from the Analysis Function and Science and Engineering Professions as well.⁸⁰

Recommendation 4: The Government Data & Al Service should be tasked with publishing guidance on Al adoption.

Within the Service, there should be:

- A team of specialists who serve as the point of contact with each department, understand its business and pipeline of projects, and advise on the spend control process.
- A central team which collects evaluations of AI from across government, and assures commonly used off-the-shelf and open source AI products on behalf of central government, making these resources available to government departments and public bodies.

⁷⁹ Lane Fox, *Directgov 2010 and Beyond: Revolution Not Evolution*.

⁸⁰ Cabinet Office, 'Government Professions', Web Page, 16 February 2024.

3. Funding

3.1 Realistically assessing the prize

Government investment in AI should be proportionate to the scale of the opportunity, but in practice the right level is challenging to judge. There is a temptation to overstate the opportunity given the growth rate of frontier technology, but an assessment of the benefits needs to be realistic.⁸¹ As Professor Karen Yeung argued, application by the State must not "rest on a naïve belief that the adoption of AI is valuable in and of itself, without first identifying and evaluating the specific concrete benefits that adoption of these systems can be expected to deliver".⁸²

Conversely, many feel that there are large savings opportunities, but simpler forms of digital technology could realise them, with one interviewee suggesting that the inclusion of AI was too often "putting lipstick on a pig". This is likely a symptom of immature uses of commonplace generative AI tasks, rather than the kind of targeted deployments this paper recommends.

There are a wide range of estimates for the benefits of full-scale AI adoption in the State. At this early stage in adoption, it would be unwise to anchor expectations to a specific overall figure. Instead, the business case should be built for investment in specific activities. Nonetheless, the scale of the estimates illustrates that the potential prize for government is large.

Public First estimates that AI adoption could provide as much as £12 billion in savings from the public sector by 2030, and up to £17 billion by 2035.⁸³ This is based on using a generative AI model to classify the probability that LLMs could significantly reduce the time it takes for workers to complete 17,000 combinations of tasks, then aggregating these at occupation level.

Based on a similar methodology, the National Audit Office referenced internal CDDO analysis which suggested almost one third of tasks in the civil service could be automated.⁸⁴ Recent research built on this, applying that estimate to the overall £175 billion of public expenditure on workforces (both the civil service, and wider public sector) to assess that the productivity benefits could be as high as £40 billion a year after discounting investment, and that in DWP specifically as much as 40 per cent of tasks could be automated.⁸⁵

Whilst serving as the Chief Operating Officer for the Civil Service, in a speech to *Reform*, Alex Chisholm estimated that AI and automation could save up to £4.8 billion a year, by reducing government's operational delivery costs by 31 per cent.⁸⁶ Then outgoing Government Chief People Officer, Rupert McNeil, told the Public Administration and Constitutional Affairs Committee that "by the mid to late 2030s, the civil service should be about 150,000 people",

⁸¹ OpenAI, 'GPT4o- Demo', Web Page, May 2024.

Karen Yeung, 'Written Evidence to the Public Accounts Committee: Use of Artificial Intelligence in Government', Web Page, May 2024. Yeung.

⁸³ Microsoft and Public First, *Unlocking the UK's AI Potential*...

⁸⁴ National Audit Office, Use of Artificial Intelligence in Government.

 ⁸⁵Iosad, Railton, and Westgarth, *Governing in the Age of AI: A New Model to Transform the State*.
⁸⁶ Sam Trendall, 'AI Could Save £5bn a Year and Take Workload from "Tens of Thousands" of Jobs, Government Ops Chief Says', *Public Technology* (blog), November 2023.

taking into account the opportunities of AI and automation – a reduction of nearly 70 per cent compared to 2023.⁸⁷

Estimates from within government are similarly significant. Dr Laura Gilbert, Director of Data Science at 10 Downing Street, and Director of the Incubator for AI, told the *Financial Times* that "whilst there was a general target for a 3.55 times return on investment on AI… "some of the early tools are more like a 200 times return on investment".⁸⁸

However, interviewees expressed scepticism about the value of many of these high-level estimates, which were felt to be unrealistic. Indeed, the benefits which central government hopes to realise from current investments are much less ambitious. The previous Government's announcements of a public sector productivity drive in the 2024 Spring Budget relied on AI and automation for much of the costed returns, expecting to realise £1.8 billion of estimated benefits by 2029 for a cost of £800 million, a more modest 2.25 times return on investment (ROI).⁸⁹

Much of the scepticism of the public servants interviewed was based on querying the State's capacity to actually deliver the benefits, with some noting that, if the estimates are less ambitious, the cost-benefit analysis may no longer stand up. As one official reflected "if you can replace 10 per cent of a team with AI, but you need to hire more civil servants with expensive tech skills to build and run it, plus more software overhead, so will it really save money overall?"

Several interviewees singled out health, the largest frontline service, as an area where scaling AI was hard because the system had not established the value of many technologies. One interviewee said "not many companies have actually done health economic studies which assess the benefits of the technology. This is where most companies [selling into the NHS] fail". Another summarised it this way:

The health system is filled with proofs of concept. But we've proven the concept, we know you can put predictive tools in the hands of clinicians and managers, and they can perform better. What we need is proof of value, and to do that we need longer studies and more investment in evaluating them.

Clearly, within government some areas will be much more amenable to AI adoption than others, and these are the areas likely to deliver best value for money. Instead of building the case for investment on broad estimates across the public sector, HM Treasury (HMT) should consider the cost of investment on specific use cases where the technology is already proven, and focus investment on prototyping and rapidly scaling up adoption.⁹⁰ To give three concrete examples:

• The European Commission's translation unit has decreased in size by 17 per cent over the last decade as a result of their use of machine translation systems⁹¹

⁸⁷ Rupert McNeil, *Oral Evidence: Civil Service Human Resources, HC 1399* (Public Administration and Constitutional Affairs Committee, 2023).

⁸⁸ Lucy Fisher, 'UK Government to Trial "red Box" AI Tools to Improve Ministerial Efficiency', *Financial Times*, 28 February 2024.

HM Treasury, '£1.8 Billion Benefits through Public Sector Productivity Drive', 2 March 2024. HM Treasury.

⁹¹ Gregorio Sorgi, 'EU Gives More Power to AI Translation Machines', *Politicio*, 15 June 2023.

- Klarna's chatbot is estimated to be performing the work of 700 full time equivalent employees in answering customer queries about payments and online financial services.⁹²
- Developers using Github Copilot to develop new code in other governments are already estimated to work 55 per cent faster.⁹³

This approach does not require bullish assumptions about future technological innovation, instead rooting the case for investment in practical case studies of the technologies already applied in the public, and comparable parts of the private, sector.

3.2 State of play

It is impossible to create a complete bottom-up assessment of current government spending on AI, because this spending is so disaggregated. But there are patterns in the parts of the system where data is available.

Headline investments in the public sector to build AI capability are either focused on evaluating foundational models through the £100 million of initial investment in the AI Safety Institute,⁹⁴ or the early capabilities to pilot and scale new tools through the Incubator for AI funded with £100 million over five years.⁹⁵ But government AI spending is likely to be much larger than this – it is just hidden in overall digital spending contracts, which are rarely broken down by the kinds of software used.

Data obtained from Tussell's procurement data platform (Figures 2, 3 and 4, full data in Appendix 2) enables a partial breakdown. Overall IT procurement, which will include outsourced AI development, has been slowly increasing for several years, rising to £13.6 billion in 2023, from £8.1 billion in 2017. Central government dominates the market, spending \pounds 7.9 billion, with health services spending £3.5 billion and local government £2.2 billion. Much of this cost will not be new spend, with plenty going to servicing legacy IT projects and programmes which are outdated and expensive to run, but also expensive to remediate and replace.

Within overall digital spending, another indicator which is important for AI adoption is cloud services. It is significantly easier to train, develop and deploy AI models in cloud infrastructure, on data held there, and to integrate those models into wider digital services hosted in the cloud. The final quarter of 2023 saw three of the larger operational government departments – the Home Office, DWP and HMRC – all award cloud services contracts to Amazon Web Services (AWS), meaning overall spending in 2023 was significantly higher than previous years – $\pounds 2.1$ billion compared to $\pounds 1.5$ billion in 2022.

It is particularly difficult to identify total spending on procuring AI technology specifically. Notably, the value of contracts which mention AI or associated keywords in their documentation appears to be decreasing. After an increase in late 2020 and throughout 2021, spending has fallen off – in 2023 contracts awarded were valued at £172 million, compared to

⁹²Jack Kelly, 'Klarna's Al Assistant Is Doing The Job Of 700 Workers, Company Says', *Forbes*, 4 March 2024.

 ⁹³Andrew Weiss, 'Empowering Developers in Government with GitHub Copilot', Github, n.d.
⁹⁴Department for Science, Innovation and Technology, 'Global Leaders Agree to Launch First International Network of AI Safety Institutes to Boost Cooperation of AI', 21 May 2024.
⁹⁵Cabinet Office, 'Whitehall Set to Bring in AI and Data Experts under Plans to Turbocharge

£358 million (more than double) in 2021. Although this increase coincides with the Covid-19 pandemic, the four largest contracts awarded in 2021 do not mention Covid-19, and so growth cannot be directly correlated. However the sample size of figures that support this market is fairly low – 806 contracts in total between 2017 and 2023 – making it harder to estimate the actual value in any given year.

Nonetheless, despite reduced spending, the number of contracts awarded has increased – from 180 contracts in 2021 to 201 in 2023. This inversion has contributed to a fall in the average contract value, from £248,000 to £126,000. There has also been a reduction in the median contract duration from 17 months to 11 months.

Figure 2: Total Awarded AI value by Central Government, Local Government and NHS in England, 2017 to 2023



Figure 3: Total Awarded AI volume by Central Government, Local Government and NHS in England, 2017 to 2023





Figure 4: Median awarded contract value for AI services, by Central Government in England, 2017 to 2023

This appears to support reports from interviewees that the majority of government investment is going into pilots, and not into scaling up products. Crucially, interviewees felt this was a function of internal public sector bureaucracy, rather than a failing of individual pilots. One described the process as "companies win a tender to pilot software, it works well, the buyer in government says 'thank you very much, we'll let you know when we next put out a call', and they never hear back". The specific challenges with procurement approaches like this are covered in Chapter 4, but the failure to scale from pilots has both cultural and financial causes.

One interviewee explained that the incentive in government are to "pilot AI in your team because it makes you cool, not because it'll do anything". Others claimed it was a symptom of digital funding being squeezed overall, and so "the problem is not what to start, but what to stop, and therefore what you can afford to scale up. Proofs of concept are cheap, financially and politically. Stopping [other programmes to scale up a pilot] is costly."⁹⁶ One department interviewee mentioned how in previous Spending Reviews, GDS had played an important role in ensuring that departments' bids properly prioritised legacy IT remediation and other crucial enabling investments, a practice which should be continued to ensure investment is not misallocated to AI without the right infrastructure to deploy it.

Overall there was broad consensus that constraint on public finances had reduced the scope to innovate and then scale up, and as a result many pilots were commissioned with very little prospect of being scaled even if they were to prove their value. This is an example of the

⁹⁶ Two separate interviewees believed too many pilots were funded from underspends in the final quarter of the financial year, although this did not match the data obtained by Tussell, which instead shows the most value being awarded in Q2 of the calendar year, i.e. the first quarter of the government's financial year.

"initiativitis" which Dame Louise Casey has described as not changing anything, but existing because it "makes governments feel better".⁹⁷

This paralysis is evident in the numbers: of the public bodies using AI in government which the NAO surveyed, only 37 per cent have one or more use case fully deployed – the rest were either piloting (25 per cent) or at an even earlier stage.⁹⁸

The status quo of many small projects, and no obvious route to scale AI in the public sector, is unsatisfactory. One interviewee said that public services should be imploring the government to "end this sea of random pilots". Another was sceptical that there was not significant duplication among the existing pilots, calling for a "systemic study of pilots funded by the government, to understand what we have and haven't learned". Steps have been taken in this direction, for example a cross-government group established to share learning on a pilot of a generative AI tool, however there is currently no systematised process for disseminating knowledge on AI opportunities within government.⁹⁹

The result is a 'valley of death' for AI projects in government, similar to the concept originally discussed in US defence procurement.¹⁰⁰ There is relatively plentiful investment to fund R&D activity and applied research in the form of pilots, but no clear path to transition projects into full-service deployments. And often investment is focused on the most novel and exciting new technologies, rather than more easily available 'quick wins': "some senior leaders ask for things that are impossible to deliver" while "ignoring things that could be readily delivered and easily provide benefit, but are not deemed sufficiently exciting."

This "leap-frogging" means that creating new things, such as the National Covid-19 Chest Imaging Database, is often prioritised above rolling out basic-but-essential universal services like electronic prescribing, which would deliver instant efficiencies for GPs and patients alike."¹⁰¹

3.3 Funding to scale

Focusing business transformation investment in applications which have a strong evidence base, and where implementation is likely to have the lowest technical and operational barriers to adoption, should build confidence. Yet even with this focus, public funding is insufficient to scale AI products, and leaders in departments and frontline services have no direct incentive to prioritise development over their day-to-day spending pressures.

A former official characterised the public sector approach as the opposite of how a technology company would fund software:

The government are obsessed with protecting frontline budgets, so [they] squeeze the small amount you have available for transformation every year. But in the private sector, they would do the opposite, because you can always

⁹⁷ Alona Ferber, 'Louise Casey: The Tories Are Done - and They Need to Go Away for a While', *New Statesman*, 29 November 2023.

⁹⁸ National Audit Office, Use of Artificial Intelligence in Government.

⁹⁹ National Audit Office.

¹⁰⁰ Eric Lofgren, 'Explaining the Valley of Death in Defense Technology', *Acquisition Talk* (blog), 9 December 2019.

¹⁰¹ Natalie Bryom, Rachel Coldicutt, and Sarah Gold, *People First, Always: Delivering Better, Cheaper, More Accessible Public Services*, 2024.

control your change budget, but if you don't focus on getting your annual run costs down then they'll keep growing.

With an incomplete understanding of current spending along with varying estimates of the level of benefits, it is not straightforward to cost the additional investment the government should make to deliver widespread AI adoption in public services. However, based on the scale of current estimates for spending (low hundreds of millions) compared to the benefits available from going further (into the billions), government is clearly missing an opportunity.

Recent estimates of productivity benefits vary from £17 billion by 2035 to £40 billion a year (net of investment), though these are likely to be the upper limit possible with existing technology.¹⁰² The benefits which can be realised in the short term, from proven technology, are likely to be an order of magnitude smaller, closer to Alex Chisholm's estimate of £4.8 billion a year.¹⁰³ This is particularly true given the government's tendency towards optimism bias in scoping the benefits of a transformation project, and the costs of realising them.¹⁰⁴ Given this paper's recommendation that the Government start by focusing on applications of AI which are feasible in the short term, and have an existing evidence base, the return on investment should be in line with government's current target of an average Return on Investment of 3.55x.¹⁰⁵

The Government should establish an AI Transformation Fund, with an additional £1 billion between 2025-26 and 2027-28, to pilot and scale up adoption of AI. This is a proportionate investment to the realistic scale of benefits in the short-term. It should only be deployed in those areas where there is existing evidence of benefit, and where the path to testing and adoption in public services is identified – in other words, it should not fund experimental new kinds of AI technology which would be classified as R&D. Within these areas, the Fund can pilot new approaches if necessary, but should predominantly be used to scale up successful existing pilots into full-service solutions.

While the Fund should be focused on the use cases outlined in Chapter 1, it should be flexible for investment in whatever technology is most important to facilitate scaling up those use cases in government, including core infrastructure like cloud and hosting, and hiring additional civil servants to create more in-house technical capability.

The majority of the Fund would be a one-off cost over the three year period, and should be treated as investment. However, some of this spending will translate into ongoing run-costs – for example, staff hired to run AI services and continuously improve them, license fees for off-the-shelf products, and cloud infrastructure to host data and software. Often Capital DEL budgets are used to fund business transformation, however research shows that "some departments have challenged using capital budgets for GenAI projects on the basis that there is no 'asset' at the end of them, due to the limited use of code, with legal teams making the case that prompts do not count in the way that code does".¹⁰⁶

Particularly given the importance of building in-house capability and procuring more flexibly through off-the-shelf products, a significant proportion of the fund should be accounted for as Resource DEL – the category of spending which can be allocated to those activities. This will

 ¹⁰² Iosad, Railton, and Westgarth, *Governing in the Age of AI: A New Model to Transform the State*.
¹⁰³ Trendall, 'AI Could Save £5bn a Year and Take Workload from "Tens of Thousands" of Jobs, Government Ops Chief Says'.

¹⁰⁴ The Behavioural Insights Team, *A Review of Optimism Bias, Planning Fallacy, Sunk Cost Bias and Groupthink in Project Delivery and Organisational Decision Making* (Department for Transport, 2017). ¹⁰⁵ Fisher, 'UK Government to Trial "red Box" AI Tools to Improve Ministerial Efficiency'.

¹⁰⁶ Paradigm Junction & PUBLIC, *Buying Generative AI in Government*, 2024.

ensure that the budget classification of the Fund does not bias it's use against the kinds of AI adoption which are better value for money overall.

The AI Transformation Fund should be managed by the GDAIS, and allocated to fund projects in departments and public services which have the greatest potential value for money. Central allocation is essential to reduce the risk of duplicating efforts across government, or funding being repurposed by departments for other projects. The Fund should not be allocated entirely up-front, with the majority of funding held back to scale up successful projects. However, the default assumption should be that successful projects supported by the Fund are scaled up, rather than requiring further approvals and bids. The following section discusses the process for allocating funding.

Given that funds will only be allocated to projects with high, near-term (two to three years) Return on Investment potential, and that this will be based on evidence from existing applications, this increase in expenditure should be more than covered by the savings accrued. However, GDAIS could consider options for offsetting some of the cost, such as closing down separate and underperforming funds which support pilot-only projects, or top-slicing existing government department technology budgets to finance the Fund.

Recommendation 5: Government should announce a new AI Transformation Fund, and HM Treasury should allocate GDAIS an additional budget of £1 billion between 2025-26 and 2027-28 to finance it.

The Fund should be available for projects which already have an existing evidence base, and demonstrably high productivity-boosting potential. There should be flexibility in what kinds of spending the Fund is used for, provided they support the scaling up of AI adoption in defined use cases.

3.4 Picking winners

The usual approach of encouraging departments to proactively bid into a central fund is insufficient, not least because the incentives to bid are not there. Cat Little, then Second Permanent Secretary at HM Treasury, told the Public Accounts Committee that the Treasury was disappointed with the track record of the Shared Outcomes Fund, which only had 28 bids in SR21 despite HMT delivering training and guidance to over a thousand finance officials on how to submit one.¹⁰⁷

Interviewees said they lacked the internal capacity to build the case for investment and manage the negotiations, without certainty of securing funding, often taking several months. This is compounded by how restrictive that spending often is, with one saying "sure, the Treasury might give me the money to buy in some consultants to build a tool, but they won't fund me to hire people to administer that contract, or manage the tool in live service. So why should [department] bid when it costs us?"

Interviewees also stressed that processes based on current business cases were too slow, with one saying: "it takes us eighteen months to build a business case. Then a year-long procurement after that. By the time that's done, the technology has just moved on". The

¹⁰⁷ Public Accounts Committee, *Oral Evidence: Cross-Government Working*, HC 75 (London: The Stationery Office, 2023).

unsuitability of the business case process for digital transformation in particular is well documented, including how they lock projects into fixed and unrealistic timeframes, and are not used as dynamic and live documents.¹⁰⁸ Far from being rigorous, these processes do not sacrifice speed to improve quality – in fact often the opposite is true.

Despite the Treasury's *Guide to Developing the Project Business Case* stating that the process should be "flexible and scaleable",¹⁰⁹ interviewees said this was not the case, and most departments practiced annual reapprovals which ran into the hundreds of pages for digital programmes. One reflected that "if officials are meant to bring all major changes to the departments' portfolio board for approval, but the board only meets once every two months, that's hardly going to work. Digital change is too quick for that".

Others were more sceptical of the value of the business case process at accurately establishing value for money, its core purpose as outlined in the Treasury's Green Book. One interviewee described a bid for hundreds of millions of funding from HM Treasury for a digital transformation programme, after which the Treasury told them "you should have added another zero, then we would have taken it seriously. That's just too small for us to fund".

Another interviewee argued that the business case process has been designed for major infrastructure programmes, where "[the government] is committing hundreds of millions of pounds, and once you've started the costs to stop are very high. That's not true where I work – software is pretty cheap to build, and costs little to stop building if it's not going anywhere. Most of our suppliers are paid for the time they work on the project anyway, if you close it down they stop charging".

One interviewee reflected:

The majority [of business cases] are rubbish. They aren't feasible or real. [Officials] never set out the real costs of delivering or running the thing, because they are incentivised to make everything have a cashable return on investment to get it through the Treasury. Then, once the project has started, it can't be stopped. It just drives the wrong behaviour.

The adverse effects of this approach are well-documented. In *The Radical How*, Andrew Greenway and Tom Loosemore blamed the high levels of 'friction' in the business case approvals process for pervasive "sunk cost fallacy", where if government "spends enough money on a programme... it can take on a life of its own, regardless of whether it is the right thing to do".¹¹⁰

Given the importance of spending public money well, the business case process must be rigorous. However it must also be proportionate, and in the case of digital projects in particular, much quicker. The Treasury should therefore agree a robust allocation process with GDAIS, after which the Fund should operate on a portfolio basis akin to that of the Advanced Research and Invention Agency (ARIA) – leaving project level funding decisions to GDAIS. The business cases used by GDAIS should be simple and quick to produce and evaluate. GDAIS should reflect the recommendations made by Lord Willets in his excellent review of the DSIT Business Case process.¹¹¹

¹⁰⁸ Sally Howes and Tess Kidney Bishop, 'The Hidden Obstacles to Government Digital Transformation', October 2018.

¹⁰⁹ HM Treasury, 'Guide to Developing the Project Business Case', 2018.

¹¹⁰ Andrew Greenway and Tom Loosemore, *The Radical How* (Public Digital, Nesta, 2024).

¹¹¹ David Willetts, Independent Review of the DSIT Business Case and Approvals Process, 2024.

The AI Transformation Fund should be gradually allocated to projects over time by Programme Directors, based on opportunities they identify within government. Given the risk of departments and public services focusing on use cases which are "tinkering around the edges" (see Chapter 2.3.2) because of aversion to the operational risks of adoption (see Chapter 5.3), Programme Directors should be supported by a team to assess which opportunities are truly transformational, and to address the barriers preventing adoption in the highest value areas. These teams will need to be hybrids of digital specialists and those with expertise in business transformation.

While the Fund will focus on projects where there is a high confidence of success, AI is still an experimental technology being deployed in a challenging delivery environment, and so it is likely some projects will fail. The crucial point is that those projects which do succeed will know that there is funding available to scale up and realise the full benefits of operating as a mature AI-powered service, something which is currently lacking.

This would operate like a presumption in favour of funding successful projects. There are parallels between this and the way venture capitalists fund startups – proceeding to fund future investment rounds if the startup is successful, and abandoning it if not. The difference is that the AI Transformation Fund is targeted at projects in which the technology is already proven, so the failure rate should not be exceptionally high.

Recommendation 6: HMT should agree an overall business case for investment with GDAIS, based on the Programme Directors' aggregate understanding of the value for money case for investment in their use cases. This business case should meet the recommendations of the Willets Review of the DSIT Business Case Process.

Recommendation 7: GDAIS Programme Directors should have the flexibility to allocate funding to projects in departments and public bodies which demonstrate the potential to represent good value for money, without reverting to HMT sign off. They should identify these opportunities by scouting them from those organisations.

Recommendation 8: Within funding they commit to projects, GDAIS Programme Directors should then have the flexibility to allocate funding to whatever capabilities are needed to deliver the project. This could include additional staff in GDAIS or the department, contracted support, software licenses, data curation and remediating legacy IT.

4. Build it, buy it, rent it

4.1 Who should develop AI for government?

The government resources software development, including AI, through different mechanisms:

- 1. Building AI in-house, using civil servants or other public sector staff to build and maintain the product.
- 2. Buying in support from the private sector to build bespoke software. This is gradually shifting from a system which incentivised the entire outsourcing of software development to the private sector, to procuring consultancy support to co-develop AI with internal staff, and owning the intellectual property in government.
- 3. 'Renting' AI capabilities through commercial off-the-shelf products, or using opensource tools and paying for the integration locally.

Broadly, interviewees felt like the current balance was skewed too heavily towards using the private sector, and the public sector needed more in-house capability. Without that, they believed it would continue to be a bad customer for software developed in the private sector, and not have sufficient capacity to test and procure off-the-shelf solutions for use in government.

However, circumstances will always vary operationally, and different departments and agencies will require different combinations of resources to build AI products in their environment. Government should not be prescriptive on the right approach, and instead should make all options easier to use effectively.

4.2 Building it: in house capability

Government does not have extensive in-house capabilities to build and deploy AI software. The Digital, Data and Technology (DDaT) profession has grown by nearly 20 per cent since April 2022, but still has vacancy rates above its target of 10 per cent.¹¹² As the NAO characterised it, "there is a major digital skills shortage in the UK, and skilled digital professionals command a premium in the market, making it hard for departments to recruit". One interviewee for this paper said that "in some DDaT functions, the vacancy rates are 30 or 40 per cent, and the majority of roles that are filled are filled by contractors rather than staff". CDDO reported to the NAO in 2023 that 37 per cent of government recruitment campaigns for digital roles were unsuccessful".¹¹³

The Civil Service People Plan commits the civil service to becoming an organisation where "specialists are offered the tools and training to deepen their expertise" and to develop a new reward strategy which would "act to attract talent to the Civil Service", however limited details are available.¹¹⁴ Different functions are already permitted some pay flexibility, with the Digital,

¹¹² Jim Dunton, 'Cabinet Office Flags Digital Successes with Three-Year Roadmap', *Civil Service World*, 29 September 2023.

¹¹³ Public Accounts Committee, *Digital Transformation in Government: Addressing the Barriers to Efficiency*.

¹¹⁴ Cabinet Office, Civil Service People Plan 2024-2027, 2024,

Data and Technology Function having a separate pay framework for its six most critical roles.¹¹⁵

The salary disparities are even more pronounced in the field of AI. This skillset is in particularly high demand, and many top machine learning engineers make salaries 10 times what they would be paid in the civil service – last year OpenAI researchers were reportedly hired on over \$800,000 a year.¹¹⁶ Even in the UK, one technologist interviewed estimated that the equivalent compensation for some AI engineers in government would be £650,000. Plenty are paid even more than that – Marc Warner estimated a "top quartile engineer with five years of experience in Silicon Valley can get four or five million dollars a year in salary… you have to offer those people a lot, a really powerful mission to get them… into the public sector, and some sensible salary".¹¹⁷

As discussed in previous *Reform* research, the government can offer higher salaries, but there are many barriers to this – mainly the significant bureaucracy involved in agreeing higher pay cases, with multiple approvals to add additional 'allowances' for exception skills, which do not follow talented individuals throughout their career and must be reapplied for when jobs change.¹¹⁸ This is an unappealing prospect for candidates who can command much higher salaries, with much more certainty and no bureaucratic hoop-jumping, in the private sector.

Conversely, one interviewee said that many businesses which required less competitive digital skills actually get away with paying less than the civil service because "you'd have to pay those people even more to work in government! Developers want to write code, not sit in meetings or make Powerpoint". Another interviewee believed that technologists weren't valued in government, saying "the civil service won't accept that people with exceptional tech talent have other options… you have to value their contributions or they will leave" – citing the example of civil servants who were expected to develop AI on laptops which they were not allowed to use Python on.¹¹⁹ They acknowledged that it was often assumed that if software engineers left they were easily replaced – contrary to the evidence on how hard it is to recruit specialist skills like AI development.

Clearly, culture and the incentives for progression are badly aligned for technical expertise in the civil service. And, as also set out in previous *Reform* work, career progression within the civil service is currently often dependent upon leaving one's area of expertise or increasing the share of one's time devoted to management responsibilities.¹²⁰ This has been referred to as "a professional ceiling where they can rise no further – the upper pay bands of the civil service being largely reserved for generalists, with generalists' strengths".¹²¹ This places limits upon the time AI experts can devote to AI and contrasts poorly with private sector companies, for example Microsoft which has separate career paths for engineers who want to focus purely upon developing their technical skills, not manage people.¹²²

¹¹⁵ Joe Hill, Charlotte Pickles, and Sean Eke, *Making the grade: prioritising performance in Whitehall* (Reform, 2024).

¹¹⁶ Jo Constanz, 'OpenAl Engineers Earning \$800,000 a Year Turn Rare Skillset Into Leverage', *Yahoo Finance*, 22 November 2023.

¹¹⁷ Civic Future, 'Should the State Play More of a Role in the Economy?', YouTube, 12 July 2024 ¹¹⁸ Hill, Pickles, and Eke, *Making the grade: prioritising performance in Whitehall*.

¹¹⁹ Python is a computer programming language, which is commonly used for developing software which uses data science and machine learning techniques.

¹²⁰ Hill, Pickles, and Eke, *Making the grade: prioritising performance in Whitehall*.

¹²¹ Greenway and Loosemore, *The Radical How*.

¹²² Chris Walden, 'How Individual Contributors Can Become Brilliant Technical Leaders', 24 June 2020.

It is unrealistic to benchmark public sector pay to the heights that exceptional skills could command in the private sector, but "there is only so far you can stretch the elastic", as John Kingman commented.¹²³ One interviewee reported that they had benchmarked what they offered in the civil service at about 30 per cent less than equivalent roles in the private sector, and that seemed to be successful at attracting some people who were highly motivated by public service.

Encouragingly, interviewees said that elements of government had made progress on this, and that the Incubator for AI had previously been able to agree higher salaries for recruits with deep technical skills. The AI Safety Institute has attracted a team with impressive technical credentials into government, and offers salaries of up to £135,000 for engineers to join – salaries which would be exceptional for many Senior Civil Servants.¹²⁴

These approaches should be replicated much more widely in government. Building on *Reform's* recommendations in *Making the grade,* government should expand the approach to pay used by the AI Safety Institute to similar technical roles in GDAIS, and put exceptional talent on a Specialist Development Scheme where their pay and progression is overseen on a case by case basis by the functional leader in their department.¹²⁵

This model is likely to be considerably cheaper overall than continuing the current dependence on private sector help, given the considerable mark-up which private businesses have to charge on top of the salaries they pay. One interviewee noted this was particularly true of "super agencies" which hired ex-civil servants, then contracted them back into government, full-time, at a significant markup to pay them more and cover business overheads.

Recommendation 9: GDAIS should establish a cross-government Specialist Development Scheme for key AI roles, as outlined in the *Reform* paper *Making the grade*. To attract candidates to join the civil service on these terms, GDAIS should pre-agree pay frameworks for crucial AI roles with the Cabinet Office, which are at least in line with the levels paid for the AI Safety Institute and Incubator for AI.

4.3 Buy it: better partnerships with the private sector

The private sector is still likely to play a key role in capacity and expertise for scaling AI across government. Indeed, a public sector with more in-house capability should be a more intelligent customer of the private sector, preventing the endless cycle of projects being procured, piloted then stopped and eventually restarted. Many of the failings of procurement come from poor decision-making about the value of different kinds of AI, based on insufficient internal understanding.

Most government digital spending is with large IT companies, who develop and manage enterprise software for government. However, the total percentage of the market the top 10 suppliers own has fallen from 68 per cent in 2017 to 45 per cent in 2023, opening up more room for suppliers and other SMEs to provide more specialist services to government, many of which have specialist capability in AI.

¹²³ Institute for Government, *Why Is Civil Service Reform so Hard? Sir John Kingman in Conversation with Bronwen Maddox*, 2020.

¹²⁴ Al Safety Institute, 'Careers', Web Page, n.d.

¹²⁵ Hill, Pickles, and Eke, *Making the grade: Prioritising Performance in Whitehall*.



Figure 5: IT procurement spend by Central Government in England, by supplier type, 2017 to 2023

Since 2011, the Government has moved away from big, end-to-end contracts to build software tools. Now, it uses a more diverse range of suppliers, requiring intellectual property rights for software to be owned by the government, and co-producing technology with civil servants – all enabled through procurement approaches like the Digital Marketplace.¹²⁶

However, because of the lack of internal government capacity to run software, several interviewees discussed how private sector expertise was still often misused. One mentioned how there were "whole teams run by contractors or companies delivering managed service contracts but for BAU [business-as-usual] processes, just because we didn't have funding certainty year to year to be allowed to hire civil servants. That's a very costly way to run those services". Another said these problems were even more pronounced in local government, where there is a real shortage of funding for digital change, meaning many areas are unable to fund any service transformation. Instead, "lots of the data about people, revenue, payments and housing is all held in bespoke, private sector services, which prevents it being brought in house... decades long outsourcing has hollowed out any local technology capacity".

A former official mentioned their endless frustration that "after a company has come in and built a product, the civil service has nobody to take it on, meaning it is never used and the same software has to be built again years later". And whilst more flexible, consultancy-based procurements make it easier to avoid supplier lock-in, the incentives created by the

¹²⁶ Government Digital Service, 'Buying and Selling on the Digital Marketplace', November 2020.

commonplace "time and materials"¹²⁷ contracts are for suppliers to take longer to produce software so they can bill more to the buyer, and to 'upsell' them on solutions they do not need.

The key challenge to more efficient procurement seems to be that large, monolithic contracts are still hard to get into, and hard to get out of. This makes the market much harder to enter for disruptive SMEs with unusual products, who do not have the 'runway' (budget) to survive long procurement processes. It also places a lot of risk on the initial contracting decision – if a public authority picks the wrong supplier, and they do a bad job, then exiting the contract and tendering a new one could take 12 to 18 months, causing the project to start over. Andrew Greenway and Tom Loosemore are right that "we need procurement systems that support nimbler, more open marketplaces,"¹²⁸ which would be easier for private companies to enter, and easier for government clients to exit if needed. Procurement needs to be much more experimental and iterative as government buyers do not always understand precisely what they need, and will only learn by working through it collaboratively, making long-term contracts undesirable.

More accessible procurement need not come at the price of sacrificing quality – often long procurement processes undermine quality rather than ensure it. As PUBLIC outlined:

There are good reasons for which suppliers with no previous government experience are often excluded from procurements (high risk) but on a project which has low monetary value and low access to sensitive data (constrained impact) these standard operating procedures can and should safely be bypassed. By contrast, failure to override this requirement may reduce the vendor pool to larger suppliers who may have no greater expertise with the new technology and be less responsive to changing technological capabilities.¹²⁹

To address this, the Government should make use of measures in the Procurement Act, which comes into effect in October 2024. The Competitive Flexible Procedure for tendering for innovation or product development allows for adjustment of requirements, and prototyping of solutions, in the early stages of a contract – letting government pay suppliers while they conduct initial tests to see what can be achieved.¹³⁰ This should be embraced for Al development work, to allow government bodies to iteratively develop their requirements over flexible, shorter contract arrangements rather than commit to multi-year programmes which neither party understands well.

These should be made available through a new AI procurement framework, with lots targeted at innovative SMEs and larger software companies capable of scaling up products. It should incentivise joint working between multiple suppliers with different technical capabilities, and be open on a rolling basis – one interviewee mentioned that a company they worked with had looked for ways to sell AI into the NHS, found out there was only one procurement framework, and it was closed to new joiners for the next four years.

In addition, to improve suppliers understanding of the public services they want to work with pre-tender, GDAIS should establish AI Sandboxes for each Programme they are overseeing-like the local authority sandbox proposed by the London Office of Technology and Innovation (LOTI).¹³¹ These would arrange prospective supplier discussions with staff, leaders and

¹²⁹ PUBLIC, Buying Generative AI in Government.

¹²⁷ Government Digital Service, 'How to Pay for Digital Outcomes and Specialists Services', July 2016. ¹²⁸ Greenway and Loosemore, *The Radical How*.

¹³⁰ 'Procurement Bill: Competitive Flexible Procedure, How Will This Work in Practice?', *Browne Jacobson* (blog), 27 September 2023.

¹³¹ Eddie Copeland, 'It's Time We Had a Local Authority Sandbox', *LOTI* (blog), 13 January 2024.

workers, and access to digital environments where they can experiment with anonymised data and user-facing applications.

Recommendation 10: GDAIS should create a new cross-government procurement framework for AI adoption in government, with different lots for supporting services. The framework should take advantage of the new Competitive Flexible Procedure to allow public bodies to bring in providers on a trial basis to test their services.

Recommendation 11: Each GDAIS Programme Director should establish an 'AI Sandbox' for the use case they lead, an environment to bring together public sector organisations and prospective partners to discuss ideas, meet prospective users and test sample data.

4.4 Rent it: using off-the-shelf and open-source products

However, the distinction of building AI in-house or procuring a bespoke product from the private sector is an outdated model, which no longer reflects how most software is bought in the private sector. The rise of Software as a Service (SaaS), available on demand for businesses through licenses, is one of the most acute transformations of the tech sector – "subscription businesses grew more than 300 per cent from 2012-2018, about five times faster than the revenues of S&P-500 companies".¹³²

Interviewees asked about this approach were universally supportive of more procurement of commercial off-the-shelf (COTS) software tools. These can be either generalist products powered by AI, such as Microsoft Copilot, or bespoke products integrating AI for the private sector, such as Beam's Magic Notes tool, to improve the productivity of social workers through note-taking and assessments.¹³³ Many private products already exist to build AI chatbots which navigate organisational policies and data to provide user interfaces to query it. And off-the-shelf solutions already exist for translation and transcriptions software, or copilot tools.

One interviewee said that "[civil servants] can't go to commercial providers and say I want to add in other features of the software I already procured from them, I have to go out to tender for a whole new bid and invite them to submit – but that takes so much longer". Another thought this was disproportionate to the cost involved in licensing software, saying a 'seat' (license) for a particular application could cost thousands of pounds a year, but building the same product in-house would cost at least £2 million – yet still procurement processes favour the latter option. Overall, they felt this was a symptom of a public sector which "thought it was special, and needs bespoke AI tools, when actually 80 per cent of what we need is the same as the private sector."

There are many advantages to buying off-the-shelf software via a license fee, rather than building a custom product in-house:

1. The up-front development cost is paid for by the provider, with greater economies of scale in doing so because they can sell that product to clients outside of a single government institution.

¹³² Frank Cespedes and Jacco van der Kooij, 'The Rebirth of Software as a Service', *Havard Business Review*, 18 April 2023.

¹³³ Beam, 'Transform Social Workers' Productivity with Al', Beam, n.d.

- 2. Licenses are flexible, can be trialled quickly, and if they are no longer good value can be exited from.
- 3. Technical support and updates are provided as part of the license (to varying levels of service), rather than needing to be resourced by the public body procuring it.

However one of the downsides, which is particularly relevant for AI products, is the expectation that these will all be tested and validated in each individual organisation. The advent of ChatGPT and other foundational AI models, accessible via the internet or enterprise software (e.g. Microsoft Office), has meant many public servants are already using AI independently in their work – probably to a much greater extent than we currently understand.¹³⁴ The Government has urged departments to be cautious about using it, but has not ruled out the use of AI models entirely.¹³⁵

David Eaves has likened this to a 'Dropbox moment' – when previous generations of digital technology like Dropbox and Wikipedia became available, and public servants began to use it without central guidance and monitoring.¹³⁶ Whilst experimentation with off-the-shelf products may yield benefits, it can incur big risks. To deploy AI in the most challenging areas of government, rigorous evaluation of off-the-shelf products will be essential, and the government will need to take a standardised approach to ensure the quality of services powered by AI is consistent with the acceptable levels – as the National Audit Office have previously recommended.¹³⁷

In the context of health in particular, this is crucial – one interviewee who had tried partnering with the NHS said that "[NHS] trusts have plenty of products being offered to them by private companies, including many which are approved by the MHRA, but which do not yet have an evidence base as economically effective. But [the trusts] don't have the internal capacity to pilot every single one of them, and across the NHS it isn't affordable to evaluate the same software repeatedly for each trust".

A good mechanism for evaluation of off-the-shelf products is mandating audits, including thirdparty audits, for products using foundational AI systems in public services.¹³⁸ Conducting these audits centrally would be much more efficient, reduce the barriers to using products for frontline services, and also reduce the risk that AI is used in circumstances where it would perform badly or incur substantial risks.

The government should take a much more flexible approach to procuring software licenses, with a specific commercial framework for procuring AI products and enabling software through a license fee. This should allow for a long period of flexible piloting and evaluation before requiring a full competitive procedure to buy enterprise-wide software. And GDAIS should provide central capacity to test and evaluate COTS tools, 'kitemarking' them for appropriate use in the public sector, based on that central evaluation and audit, so that frontline services can take more informed decisions about which are suitable for use. Global digital standards on AI are still under development, and so GDAIS will need to develop their own policies in line

¹³⁴ Jonathan Bright et al., 'Generative AI Is Already Widespread in the Public Sector', *arXiv Preprint:* 2401.01291, 3 January 2024.

¹³⁵ Ben Gartside, 'ChatGPT: Civil Servants Warned Not to Use AI Chatbots to Write Policies and Carry out Government Work', *iNews*, 21 February 2023.

¹³⁶ Matt Davies and Elliot Jones, *Foundation Models in the Public Sector* (Ada Lovelace Institute, 2023).

¹³⁷ National Audit Office, Use of Artificial Intelligence in Government.

¹³⁸ Davies and Jones, *Foundation Models in the Public Sector*.

with emerging standards as much as possible, but should not let the absence of existing ones hold them back from pushing forward.

A similar model to this is also relevant for the open-source tools. Meta have produced and open-sourced the codebase for Llama 3.1, their most powerful language model, which on some benchmarks performs as well as models produced privately by OpenAI, Google and Anthropic.¹³⁹ Contrary to many expectations, open-source software is often considered more secure than proprietary software developed by private communities, because of the large networks who test it and suggest updates and bug fixes.¹⁴⁰ Open source products already underpin large parts of privately developed software, and plenty are integrated into government technology stacks. However, relying on open-source products to be 'lifted and shifted' into the public sector is inherently risky without adequate internal capability to integrate and evaluate them and monitor their performance on a continuous basis.

Recommendation 12: GDAIS should develop and run a new procurement framework exclusively designed for the procurement of off-the-shelf AI tools and enabling infrastructure, with an extremely light-touch initial process for trialling and experimenting with new products before going out to a full competitive process. This should be in recognition that the financial commitments required to experiment with products available via license fee are inherently much lower than other kinds of procurement, so less risky.

Recommendation 13: GDAIS's central guidance and support function should assess offthe-shelf and open source AI products and enabling infrastructure, and provide a list of approved ('kitemarked') products with supporting assessments to public bodies to support their procurement decisions.

 ¹³⁹ Meta, 'Introducing Llama 3.1: Our Most Capable Models to Date', *Meta AI* (blog), 23 July 2024.
¹⁴⁰ Dale Murray, 'Open Source and Security: Why Transparency Now Equals Strength', *Network Security* 2020, no. 7.

5. Development and deployment

Al is being developed and deployed throughout Government. A recent NAO report found 74 reported use cases, as of Autumn 2023.¹⁴¹

The Department for Transport is using AI to detect fraud through image analysis and to improve their response time and accuracy in relation to public consultation responses; DHSC is using AI to support operational decision-making such as prioritisation, eligibility and enforcement; HMRC is using AI to identify non-compliance in the tax system and help customers find the information they are looking for; HMT is using AI to triage inbound correspondence; MoJ was using AI as a short-term custody predictor and to support Actuarial Risk Assessment Instruments (ARAIs) and short-term prison demand modelling; and NHSE is using AI to moderate user reviews on NHS.uk and to provide a three-week forecast of expected A&E admissions.¹⁴²

Nevertheless, the speed with which these use cases are being trialled and, if successful, scaled up is being hampered by issues to do with policies, data, guidance and evaluation. One interviewee summed up the problem as being one of speed – "it can take months to get clearances through, weeks to get someone on-site once they have clearance, and then six months to share the data they need with them. It feels like people are constantly throwing hurdles in your way".

5.1 A principle-based approach

To develop and deploy AI, guidance needs to exist on how to do this effectively.

To a degree, guidance does exist. The *Generative AI Framework for HM Government* provides guidance on using generative AI safely and securely for civil servants and people working in government organisations, though it does not cover other types of AI.¹⁴³ The Algorithmic Transparency Recording Standard (ATRS) provides a standardised way for public sector organisations to record and share how they are using algorithmic tools, although the Department for Science, Innovation and Technology (DSIT) only announced an intention to make it mandatory for government departments in February 2024.¹⁴⁴

However, this guidance is simultaneously too much and too little to effectively encourage the development and deployment of AI.

The AI guidance which exists is voluminous and complex, and some government bodies interviewed by the NAO described finding it difficult to navigate.¹⁴⁵ The *Generative AI Framework for HM Government* runs to 74 pages and references multiple other documents that those deploying AI should consider.¹⁴⁶ Given each department and public body will have

¹⁴¹ National Audit Office, *Use of Artificial Intelligence in Government*.

¹⁴² See Appendix 3

¹⁴³ HM Government and Central Digital and Data Office, *Generative AI Framework for HM Government*.

¹⁴⁴ National Audit Office, *Use of Artificial Intelligence in Government*.

¹⁴⁵ National Audit Office.

¹⁴⁶ HM Government and Central Digital and Data Office, *Generative AI Framework for HM Government*.

their own guidance about digital adoption and frontline public service delivery, complex central guidance is not conducive to supporting scaling AI development and deployment. As one interviewee told us, "any [civil servants] who want to try and innovate will always be worried that we missed something in a footnote of a document we never saw, and that's what we get blamed for if it goes wrong".

A report by The Alan Turing Institute supported this, finding a low awareness among civil servants of the guidance on the appropriate use of Al.¹⁴⁷ The report questioned whether "professionals are even aware this guidance exists, or the extent to which it is useful in their day-to-day engagements with the technology".¹⁴⁸

On the other hand, despite the length of the guidance, few specifics are provided. For example, it is unclear at what stage of the deployment process — be that initial internal pilot or widespread roll-out — different policies should apply. The lack of clarity about requirements is likely to result in most civil servants assuming they need to comply with all of it, all of the time.

Clarity is not provided locally in departments or public bodies either. In response to an NAO survey only 21 per cent of surveyed government bodies reported having a strategy for AI and only eight said that they are always or usually compliant with the ATRS. Of the nine departments which responded to an FOI request for this paper only HMRC reported having AI guidance specific to their department, although MoJ are designing new assurance checks.¹⁴⁹ One interviewee attributed this to a "lack of clarity on what you should do. Everyone thinks someone else is going to come up with a straightforward process to follow, but what you get from the [Cabinet Office, including CDDO and GDS] isn't helpful".

This results in slow AI scaling, friction and civil servants being discouraged from adopting AI. Even in very low-risk settings, departments are forced to spend an excessive amount of time considering guidance — or asking colleagues for advice in the absence of guidance — before deploying AI. Furthermore, given the difficulties and risks associated with AI, many may conclude that it is not worth the effort. Only 37 per cent of the government bodies which responded to the NAO said that they had deployed AI¹⁵⁰ and in response to an FOI request for this paper only five departments – DfT DHSC, HMRC, HMT and NHSE – confirmed that they are still deploying the use cases which they submitted to the NAO.¹⁵¹

Standardised guidance from the centre is clearly required. However, different departments should be able to adapt guidance to their own circumstances and use cases. For example, the risk tolerance and legal obligations for deploying AI in the Ministry of Defence will be different to those in the Ministry of Housing, Communities and Local Government. Guidance should be flexible enough to allow for this. As one interviewee put it, "I need to take a different approach to assessing the risks and opportunities in [my department] than other departments would, and [I need] the flexibility to do that".

Thus, the guidance provided from the centre should be short, clear and principles-based, with departments using this guidance to develop their own guidance specific to their circumstances.

This approach of having centrally-provided principles corroborated by more specific guidance in departments is utilised in other areas of government policy. For example, the *Environmental*

¹⁴⁷ Alan Turing Institute, *Generative AI Is Already Widespread in the Public Sector*, 2024.

¹⁴⁸ Ibid.

¹⁴⁹ See Appendix 3

¹⁵⁰ National Audit Office, Use of Artificial Intelligence in Government.

¹⁵¹ See Appendix 3

Principles Policy Statement aims to provide "ministers, and those developing policy on their behalf, with the space to use the principles to enable and encourage innovation"¹⁵² and *The Civil Service Code* details how "individual departments may also have their own separate mission and values statements based on the core values".¹⁵³

This approach to the development and deployment of AI has also been used in other countries, for example in the USA, executive departments and agencies are required to use AI in accordance with eight principles, running to just under 1,300 words, while also "taking into account the views of other agencies, industry, members of academia, civil society, labor [sic] unions, international allies and partners, and other relevant organizations".¹⁵⁴

Figure 6: The White House's *Executive Order on the Safe, Secure and Trustworthy Development and Use of Artificial Intelligence*

The US Federal Government's *Executive Order on the Safe, Secure and Trustworthy Development and Use of Artificial Intelligence* sets out eight guiding principles and priorities which executive departments and agencies should adhere to. Abridged versions of these eight principles are listed below.

- 1) Artificial Intelligence must be safe and secure.
- 2) Promoting responsible innovation, competition, and collaboration will allow the United States to lead in AI and unlock the technology's potential to solve some of society's most difficult challenges. This effort requires investments in AI-related education, training, development, research, and capacity, while simultaneously tackling novel intellectual property (IP) questions and other problems to protect inventors and creators.
- 3) The responsible development and use of AI require a commitment to supporting American workers.
- 4) Artificial Intelligence policies must be consistent with the Administration's dedication to advancing equity and civil rights.
- 5) The interests of Americans who increasingly use, interact with, or purchase AI and AI-enabled products in their daily lives must be protected.
- 6) Americans' privacy and civil liberties must be protected as AI continues advancing.
- 7) It is important to manage the risks from the Federal Government's own use of AI and increase its internal capacity to regulate, govern, and support responsible use of AI to deliver better results for Americans.
- 8) The Federal Government should lead the way to global societal, economic and technological progress, as the United States has in previous eras of disruptive innovation and change.

Source: The White House, *Executive Order on the Safe, Secure and Trustworthy Development and Use of Artificial Intelligence*, Webpage, 2024.

¹⁵² Department for Environment Food & Rural Affairs, *Environmental Principles Policy Statement*, 2023.

¹⁵³ Civil Service, 'The Civil Service Code', Webpage, 16 March 2015.

¹⁵⁴ The White House, 'Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence', Web Page, 30 October 2023.

Recommendation 14: GDAIS should produce and publish a single, simple set of principles for using AI in public services.

Recommendation 15: Departments and public bodies should publish their own policies, applying these central principles to their own specific circumstances.

5.2 Data quality and access

Data is needed to train AI models, run them in deployed settings, and test and evaluate their performance on an ongoing basis. High-quality data allows AI systems to learn effectively, increasing accuracy and decreasing the risk of biases.

Government data is often of a poor quality and insufficient for training AI. And within government, that data can be technically challenging to access, especially to access from the kind of data science environments necessary to train AI tools – which are predominantly cloud based. Often, data collected may be missing key information points which are crucial for training AI. This reduces the ways in which the data can be used generally and in AI specifically. One example of this is the Windrush scandal, where data about individuals' status was not complete enough to identify those who had a right to live in the UK.¹⁵⁵

There are also multiple different ways in which data is recorded. In 2019 the NAO found more than 20 different identifiers being used to identify individuals and businesses across ten government departments and agencies.¹⁵⁶ Large public databases often do not have standard formats for capturing dates of birth or naming conventions. This means that these are inconsistently recorded. Indeed, sometimes data models and standards are different not only between departments but also within departments.¹⁵⁷ This also reduces the government's ability to use that data, because repeatable patterns for data ingest and processing cannot be reused across all development processes. The Number 10 Data Science (10DS) team have published the code to share data with them on GitHub, providing one avenue to simplify and standardise data sharing processes.¹⁵⁸

Raven Sentry was a project by the US armed forces to use AI to provide advanced warning of insurgent attacks in Afghanistan between 2019 and 2020. Analysis published by the US Army War College shows that addressing data quality and access issues was crucial in developing the programme:

Raven Sentry's development was only overcome by limiting the algorithm's geographic focus and dedicating considerable time to data curation early on... Difficulty with data formats, particularly when attempting to ingest a variety of information, is a regular theme of AI application studies.¹⁵⁹

¹⁵⁵ National Audit Office, *Challenges in Using Data Across Government*.

¹⁵⁶ National Audit Office.

¹⁵⁷ National Audit Office.

¹⁵⁸ Laura Gilbert, *Changing Government One Datastream at a Time* (Heywood Quarterly, 2024).

¹⁵⁹ Thomas W. Spahr, *Raven Sentry: Employing AI for Indications and Warnings in Afghanistan*, vol. Volume 54, Parameters (The US Army War College Quarterly, 2024).

62 per cent of the government bodies responding to the NAO identified access to good-quality data as a barrier to implementation.¹⁶⁰ One interviewee for this paper reflected: "people got quite excited about Gen AI last year but then realised our data is not quite good enough and stopped pushing", and another described attempting to develop generative models based off the government's data systems as "like trying to plug a V8 engine into an old 90s Skoda". These problems are also true of much more traditional data science.

The NAO reported that one department had to deploy 300 people to clean data in order to carry out an analysis.¹⁶¹ During the Covid-19 pandemic nearly 16,000 cases went unreported in England because Public Health England used Excel's XLS file format — which dates back to the 1980s — to pull together data.¹⁶² And one interviewee for this paper detailed how, when working on a diagnostic contract for an NHS trust, half of the time had to be spent cleaning up existing data to get it into a usable state.

In addition to the quality of the data, much government data is locked in outdated legacy IT systems. These are a significant barrier to introducing AI into the processes they support. These old systems carry significant security risks, are costly to maintain and — most importantly — they make data difficult to access. Often they were not designed with interoperability in mind or contractually the public services cannot extract the data without paying the supplier to do so.

More than 40 legacy systems are 'red-rated',¹⁶³ meaning that "the likelihood of encountering issues or failures is significant, and the potential impact of these issues could be severe".¹⁶⁴ However, the actual number of legacy systems is difficult to determine: in the past departments have found it a challenge to define and identify all of their legacy systems.¹⁶⁵ In 2019 it was estimated that nearly half of all technology expenditure across government was dedicated to "keeping the lights on activity on outdated legacy systems".¹⁶⁶ An interviewee for this paper agreed that "the scale of legacy technical debt in government is huge and makes it very difficult to use these technologies; the amount of time and money which it takes to do the data processing is vast".

To effectively develop and deploy AI, data quality and access needs to improved, and legacy systems need to be remediated and replaced.

Previous governments recognised this data quality issue and took steps to address it. The *Roadmap for digital and data, 2022 to 2025* sets out six cross-government missions to address data quality and legacy system issues, backed by £8 billion committed in the 2021 Spending Review for digital, data and technology transformation.¹⁶⁷ CDDO has also provided advice on

¹⁶⁰ National Audit Office, *Use of Artificial Intelligence in Government*.

¹⁶¹ National Audit Office, *Challenges in Using Data Across Government*.

¹⁶² Leo Kelion, 'Excel: Why Using Microsoft's Tool Caused Covid-19 Results to Be Lost', *BBC News*, 5 October 2020.

¹⁶³ Tevye Markson, 'Scale of Government Legacy IT Systems at "Critical" Risk Revealed', *Civil Service World*, 10 January 2024.

¹⁶⁴ Central Digital and Data Office, *Guidance on the Legacy IT Risk Assessment Framework*, 2023.

¹⁶⁵ National Audit Office, *Digital Transformation in Government: Addressing the Barriers to Efficiency*. ¹⁶⁶ National Audit Office.

¹⁶⁷ Central Digital and Data Office, *Transforming for a Digital Future:* 2022 to 2025 Roadmap for *Digital and Data - Original*, 2022.

coping with legacy systems, for example *Managing legacy technology*¹⁶⁸ and *Guidance on the legacy IT risk assessment framework*.¹⁶⁹

Nevertheless, over the past 25 years there have been 11 different digital strategies aimed at addressing the problems of legacy systems and data quality.¹⁷⁰ According to the NAO, none of these adequately addressed the issue and have had limited success. Previous approaches have not sufficiently addressed the underlying issues, instead prioritising 'front-end' citizen transactions rather than more 'back-end' changes, building automation on top of legacy technology.¹⁷¹

The scale of the challenge of addressing all data quality, data access and legacy system issues — and the failure of previous attempts — demonstrates that this is a major undertaking which could delay the deployment of AI. The focus, therefore, should be on improving data quality in the areas where the opportunities to use AI in the short term are most promising but are being held back by these data and legacy system issues.

This drive needs to come from the centre of government. Public servants in government departments and local government may be less incentivised to spend time and money improving data quality if it comes at the cost of investing in day-to-day priorities or if the benefits would mainly be seen in other parts of government. As the Data Advisory Board found, "return on investment for a department can often be difficult to justify in data projects because the benefits might be seen elsewhere in government".¹⁷² One interviewee for this paper said that because of competing pressures it "doesn't matter how often I say you have to put data into X format and input it in Y way... they won't do it".

Sometimes these problems are created by higher security systems, which do not have the necessary software development tools on them to build AI. One interviewee said they had been told to "write the programme in one secure environment, burn it onto CDs, and transport those by courier into another secure area" – adding that it took 400 CDs in total.

[']Fixing the plumbing' of data and legacy systems to enable AI adoption requires significant investment, but data projects have previously been set aside when funding is tight.¹⁷³ Departments often find it is easier to bid for capital funding for new digital development than resource funding to maintain existing services and it can be difficult to make the case for long-term data projects which may not provide immediate cashable benefits.¹⁷⁴ Work by CDDO suggests that only around three quarters of the 20 highest-risk legacy systems have agreed funding to complete remediation work and just over half of these plans are expected to complete their activity by 2025.¹⁷⁵

Interviewees for this paper felt that the political interest in AI had made this harder, not easier: "You can't convince ministers to spend money on data plumbing [infrastructure and quality improvements], they just want to spend it on Generative AI". Another was sceptical that the government's announcement of a new National Data Library to make this information more accessible would be successful unless it was able to go further than previous initiatives – the

¹⁶⁸ Central Digital and Data Office, *Managing Legacy Technology*, 2019.

¹⁶⁹ Central Digital and Data Office, *Guidance on the Legacy IT Risk Assessment Framework*.

¹⁷⁰ National Audit Office, *Digital Transformation in Government: Addressing the Barriers to Efficiency*. ¹⁷¹ Ibid

¹⁷² National Audit Office, *Challenges in Using Data Across Government*.

¹⁷³ Ibid.

¹⁷⁴ National Audit Office, *Digital Transformation in Government: Addressing the Barriers to Efficiency*. ¹⁷⁵ Ibid.

Data Marketplace and Integrated Data Service, the latter of which is in a beta (testing) phase before going live in 2024.¹⁷⁶

The Tony Blair Institute and Faculty's research estimated the cost of linking and cleaning government data to prepare it for implementing AI tools at between £1.25 billion and £2.5 billion.¹⁷⁷ But a prioritised approach, focused on preparing only the data which would be directly useful for the kinds of AI applications set out in Chapter 1, those which promise the most immediate return on investment, would likely cost significantly less. As outlined in Chapter 3, GDAIS Programme Directors should have the flexibility to use the AI Transformation Fund to address any technical issues which are a barrier to AI adoption, including data quality and access issues.

5.3 Risk assessment

All government transformation comes with risks. This is particularly the case with automated systems, some attempts of which have had very public failings – such as the Home Office's Streaming Algorithm, operating between 2015 and 2020, which was informed by biased data and discontinued after legal action.¹⁷⁸ And failure to identify risks can come with steep costs to the public and the State – the Australian RoboDebt system wrongly accused benefit recipients of owning the government money, culminating in them repaying debts of over \$721 million and settling a class action lawsuit for over \$1 billion.¹⁷⁹

Being conscious of risks is, therefore, crucial if major changes are to succeed. But interviewees for the paper were almost unanimous that this had gone too far and risk-aversion was now a key factor holding back the benefits of AI in public services, even in cases where it would be far less risky to use it than continue not using it. They felt this imbalance was created by the processes and bureaucracy of Whitehall, which stifled innovation. A participant in a recent *Reform* roundtable reflected on their time as a civil servant: "for every nine meetings I went to about the risks of AI, I went to one about the opportunities to use it".

One way for the State to have a more mature discussion about the risk of using AI is for the government to define the appropriate upper bound for the risk of a given application failing which it is prepared to tolerate – similar to the use of F-N curves for assessing societal-level risks.¹⁸⁰ By continuously evaluating model performance against this failure risk, including through approaches such as "gatekeeper AI", the government can produce a consensus in favour of using an "AI with quantitative safety guarantees".¹⁸¹

¹⁷⁶ Department for Science, Innovation and Technology, 'AI Expert to Lead Action Plan to Ensure UK Reaps the Benefits of Artificial Intelligence'. and Office for National Statistics, 'Integrated Data Service (IDS)', August 2024

¹⁷⁷ losad, Railton, and Westgarth, *Governing in the Age of AI: A New Model to Transform the State*.

¹⁷⁸ Anna Dent, *Automating Public Services: A Careful Approach* (Promising Trouble, 2024). ¹⁷⁹ Dent.

¹⁸⁰ Eric Madsen, 'Farmer's Diagram, or F-N Curve. Representing Society's Degree of Catastrophe Aversion', 24 July 2022, Risk Engineering, n.d.

¹⁸¹ David 'davidad' Dalrymple, Safeguarded AI: Constructing Guaranteed Safety (ARIA, 2024).

5.3.1 Legal requirements

Legal requirements around the development and deployment of AI mainly centre around the use of personal data and how decisions are made. GDPR Article 22 provides individuals with the right "not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her".¹⁸² Because of this legal imperative, having a 'human in the loop' is a foundational part of most government uses of AI.

Having a 'human in the loop' is also important for the practical reason that humans are needed to assure the decisions made by AI models to ensure they are right. This is especially the case at the current stage of AI development, as its capabilities are still being understood and well documented risks, biases and inaccuracies in AI's output remain.¹⁸³

One interviewee discussed how using AI as a "decision support tool" for human decision makers, for example by ranking the order of cases for a caseworker to prioritise but giving them the option to change it, could be an "deployment ramp" to full-scale automation in the future, by testing the performance of AI at parts of a task and gradually increasing automation as performance becomes proven.

Government departments and public bodies wishing to process data in an automated way are required to complete internal processes — for example a Data Protection Impact Assessment — to ensure that they comply with all legal requirements. Guidance is also provided, for example the Information Commissioner's Office's (ICO) *Guidance on AI and Data Protection*. And some departments have established bodies and allocated support to scrutinise AI use cases. The Department for Work and Pensions has established an advice and assurance group,¹⁸⁴ and in DBT "the Data Science team occasionally provide additional technical support where required", consisting of a data scientist consulting with the team who wishes to deploy AI.¹⁸⁵

It is vital that government departments and public bodies comply with all legal requirements when utilising AI. But a risk-averse attitude should not prevent the usage of AI, especially when human decision-making is already frequently being litigated. Between 2013 and 2022 87,646 asylum appeals were made and more than a third of these were granted.¹⁸⁶ Between October and December 2024 70 per cent of Personal Independence Payments (PIPs) appeals were granted.¹⁸⁷

In addition, automating elements of public services may be much less harmful than the status quo, often in the areas where there are already high levels of risk in the current system. One participant in a recent *Reform* roundtable gave the example of children's social services where concern has been expressed about the risks of AI, despite the risks to children of paperwork not being properly processed being much higher.

¹⁸² European Parliament and The Council of the European Union, *General Data Protection Regulation*, 2016.

¹⁸³ Ada Lovelace Institute, *AI Assurance? Assessing and Mitigating Risks Across the AI Lifecycle*, 2023.

¹⁸⁴ National Audit Office, Use of Artificial Intelligence in Government.

¹⁸⁵ See Appendix 3

¹⁸⁶ Molly Blackall, 'One Asylum Appeal Lodged Every Hour on Average in Britain', *The i*, 18 February 2024.

¹⁸⁷ Ministry of Justice, 'Tribunal Statistics Quarterly: October to December 2023', Web Page, 14 March 2024.

Furthermore, having a 'human in the loop' can mean that many services are still expensive event when automated, because significant effort needs to be expended checking that the AI is right. For example, microcars fitted with number plate recognition cameras are being trialled to detect illegal parking in Milton Keynes, Lambeth and Southwark.¹⁸⁸ Parking enforcement administration costs total an estimated £100 million a year in London alone.¹⁸⁹ However, microcars are limited in the extent to which they can reduce this cost by the requirement for ticketing itself to be carried out by an actual in-person enforcement officer.¹⁹⁰ And, as one interviewee for this paper pointed out, having a 'human in the loop' does not solve all problems: "take the Horizon IT system. There were plenty of humans in the loop. It still failed terribly".

Ultimately, the State needs to consider the counterfactuals of using AI for specific use cases. In some cases, this will be a service being carried out poorly without using AI, and mitigations to the risks posed by AI will be sufficient to make that a preferrable option for the government. To rebalance the risks of automation against many of the risks which come from continuing with business as usual, government policy and legal advice should consider the risks of not adopting AI in addition to the risks of adopting AI.

Recommendation 16: Government legal advice, internal processes and documents (including Data Protection Impact Assessments) should give equal parity to the risks of not using AI or automated processing, as they do to the risks of using AI to partially or completely automate the process. This should be based on the performance and risks of the current system and processes used.

5.3.2 Data sharing

Being able to share data is crucial to the development and deployment of AI. This is more important for some uses of AI than others, for example preventative applications in education, health and crime. These aim to use machine learning to identify cases for early interventions and require combining different data sets to see if models can be significantly more effective. The focus of this paper is on more mature applications of AI which do not always require data sharing. Instead, they are often built on top of data which is already held by a single public body. Nonetheless, in many areas, data sharing remains a barrier to effective AI exploitation.

Government has made progress since the Thomas and Walport 2008 data sharing review found that those who wish to share data face a "fog of confusion".¹⁹¹

GDPR set out principles and obligations for the processing of personal data, with departments responsible for keeping their own data secure.¹⁹² The Digital Economy Act 2017 sets out data sharing arrangements within the public sector,¹⁹³ the Data Protection Act 2018 provides a framework for data protection and the previous Government was attempting to pass the Data

¹⁸⁸ Nicholas Hellen, "Robowardens" to Scan Whole Streets for Parking Violations in Minutes', *The Sunday Times*, 27 July 2024.

¹⁸⁹ Hellen.

¹⁹⁰ Hellen.

¹⁹¹ Richard Thomas and Mark Walport, *Data Sharing Review Report*, 2008.

¹⁹² European Parliament and The Council of the European Union, *General Data Protection Regulation*.

¹⁹³ HM Government, 'Digital Economy Act 2017' (2017).

Protection and Digital Information Bill through Parliament, which aimed to create a new post-Brexit UK data rights regime.¹⁹⁴

When sharing data, government bodies produce Data Sharing Agreements (DSAs) and Data Protection Impact Assessments (DPIAs). The Information Commissioner's Office (ICO) is responsible for enforcing and regulating these requirements and in recent years it has expanded its capabilities, with the power to issue reprimands and fines to public bodies which misuse data. In 2018 the Met Police were reprimanded following several issues around the handling of criminal intelligence files.¹⁹⁵

These recent pieces of legislation and guidance have been beneficial. They have clarified data sharing and governance policy, and made data sharing processes more transparent and faster to implement.¹⁹⁶

However, given the understandable concern of public bodies about inadvertently undermining public trust, this body of legislation and guidance can encourage a risk-averse attitude towards data sharing. Some departments have expressed concerns about the safety of their data if they share it with others.¹⁹⁷ And, given the scale of fines which can be applied, some in government are prioritising "ensuring everyone was GDPR compliant, rather than looking at different and effective ways of using data".¹⁹⁸ One interviewee with experience working with government from the private sector said that they had been asked to sign up to a data sharing policy that they were not allowed to read before signing it, because the agreement had to be held on a system they needed to have signed it to access.

Balancing the risks of sharing data with the risks of not sharing data is not easy. As the *Generative AI Framework for HM Government* states in ambiguous language, a "delicate balance" needs to be struck between sharing data for training LLMs and minimising the collection and storage of personal data to meet the UK GDPR requirement of storage limitation".¹⁹⁹ Nevertheless, risk-aversion to sharing data must not be a barrier to the development and deployment of AI.

Some parts of and actions by government do recognise the risks of not sharing data. The Counter-Terrorism Strategy (CONTEST) recommends increased data sharing in order to better respond to threats and mitigate the risk of terrorism,²⁰⁰ and the response to Covid-19 relied on government guidance that the NHS could share data more freely to protect the public. Dame Fiona Caldicott, the National Data Guardian, outlined publicly how "information may need to be shared more quickly and widely across organisations than normal, or different types of information may need to be collected and used" to save lives.²⁰¹

¹⁹⁴ John Woodhouse, *Data Protection and Digital Information (No.2) Bill: Progress of the Bill* (House of Commons Library, 2023).

 ¹⁹⁵ Information Commissioner's Office, 'ICO Issues Reprimand to the Metropolitan Police Service for Inadequate Handling of Files Related to Organised Crime Groups', Web Page, 16 March 2023.
¹⁹⁶ Gavin Freeguard and Paul Shepley, *Legislation to Support Data Sharing* (Institute for Government, 2022).

¹⁹⁷ National Audit Office, *Challenges in Using Data Across Government*.

¹⁹⁸ Freeguard and Shepley, *Legislation to Support Data Sharing*.

¹⁹⁹ HM Government and Central Digital and Data Office, *Generative AI Framework for HM Government*.

²⁰⁰ HM Government, CONTEST: The United Kingdom's Strategy for Countering Terrorism, 2018.

²⁰¹ Fiona Caldicott, 'Data Sharing During This Public Health Emergency', Web Page, 3 April 2020.

Figure 7: Data sharing during the Covid-19 pandemic

In April 2020, Dame Fiona Caldicott, National Data Guardian in NHS England, wrote to leaders across NHS trusts and the social care system about the vital role that data would play in the pandemic, and how approaches to data sharing needed to change in an emergency:

"Information sharing must be done differently to support the fight against COVID-19 and to protect citizens compared to ordinary times. Information may need to be shared more quickly and widely across organisations than normal, or different types of information may need to be collected and used.

We know from dialogue with the public, conducted by our organisation and others in the past, that there is strong support for the use of health and care data where there is a clear public benefit. People are generally altruistic about the use of their data and want it to be used to help others as long as there are appropriate safeguards in place. I understand that doing things differently can usher in uncertainty about what is and isn't appropriate from a data sharing perspective; the worry may be that people will share too much or too little. So, we must make sure that we have the balance right to protect that admirable altruism."

This approach underpinned several novel national technologies, including ones which integrated machine learning software, which supported the Government's response to the pandemic.

The Covid-19 Early Warning System used data from hospitals across the country to forecast demand for admissions and life-saving equipment, allowing the Government to best target its limited resources to prevent the loss of life. The NHS App – which shared anonymised data with the UK Health Security Agency (UKHSA) – prevented one million Covid-19 cases, 44,000 hospitalisations and 9,600 deaths in its first year. On a daily basis the UK Health Security Agency updated a dashboard which displayed Covid-19 data and statistics at both a national and regional level. The use of dexamethasone to treat Covid-19 was discovered thanks to the use of data.

Source; Dame Fiona Caldicott, Data sharing during this public health emergency, 3 April 2020; Faculty, 'Covid-19: Helping the NHS to save lives by forecasting patient demand', 22 April 2021; Michelle Kendall et al., 'Epidemiological Impacts of the NHS Covid-19 App in England and Wales Throughout its First Year', Nature Communications 14 (2023); Doug Faulkner, 'NHS Data Strategy: Hancock Defences Data Sharing Plan', *BBC News*, 22 June 2021.

A similar shift towards considering the risks of not sharing data and a presumption in favour of data sharing should take place across government. This could be achieved via primary legislation and some countries have gone down this route. Six AI strategies of EU member states highlight the need to reform data sharing laws or directories.²⁰²

However, given the time required to take legislation through Parliament, the immediate focus should be upon reconfiguring internal government governance processes so that the risks of not sharing data or processing with automated technology are required to be considered as well as the risks of sharing data, and placed on an equal footing. This is not currently the case.

²⁰² Joint Research Centre, *AI Watch: European Landscape on the Use of Artificial Intelligence by the Public Sector*, 2022.

For example, of the eight government departments and bodies (DBT, DEFRA, DfE, DfT, DHSC, DSIT, DWP, MoJ) which provided templates of any DPIAs, DSAs and Memorandum of Understandings in response to an FOI request submitted for this paper, none required a consideration of the risks of not sharing data.²⁰³

Recommendation 17: Government internal processes and documents for data sharing, including Data Sharing Agreements and DPIAs should require officials to also assess the risks of not sharing data, and give these equal parity with the risks involved in any new sharing.

5.4 Evaluating AI

Just as it is important that AI is deployed first in the areas where it performs best and can deliver the greatest improvements to public services, it is essential that government has mechanisms to monitor the ongoing performance of AI to assess whether it meets the required standard.

This risk can be seen in cases like Babylon Health's health app, which the Medicines and Healthcare products Regulatory Agency (MHRA) investigated after one doctor complained that it had failed to identify symptoms of a heath attack and deep vein thrombosis.²⁰⁴

Many interviewees reflected that evaluating the performance of AI was challenging due to the lack of understanding of the performance of public services before AI was tested: the 'baseline' of human performance. One said that "before GDS began working with the top 75 [digital] services, most [departments] could not tell you how much they cost to run, how many people they served, and how well they performed". This starting point is essential to fairly evaluate the performance of AI against other opportunities to improve public services.

Information also needs to be available about what impact AI has upon performance. As interviewees for this paper said, government needs "to talk about the actual outcomes when monitoring" and "measuring success is key". The *National AI Strategy – AI Action Plan²⁰⁵* contains no outcome measures and the *Generative AI framework for HM Government* is vague on how to carry out evaluation, stating that the model's outputs should be evaluated against "ground truth or expert judgement, and obtain user feedback to understand the usefulness of the returned response".²⁰⁶

Numerous interviewees commented that CDDO should play a larger role in establishing baselines and evaluating ongoing performance. One said that "there is a role for CDDO here, where they should evaluate [commercial] products and kitemark" those which were high-performing, principally in the healthcare sector. Another compared the task to that performed by the Evaluation Taskforce (ETF), saying that "I would like to see CDDO take a stronger role in assuring digital and data spend, in the same way that the ETF are doing for evaluation spend".

²⁰⁶ HM Government and Central Digital and Data Office, *Generative AI Framework for HM Government*.

²⁰³ See Appendix 3

²⁰⁴ Aliya Ram and Sarah Neville, 'High-Profile Health App Under Scrutiny After Doctors' Complaints', *Financial Times*, 13 July 2018.

²⁰⁵ Department for Science, Innovation and Technology, *National AI Strategy - AI Action Plan*, 2022.

In the absence of good evaluations of current services – only a small proportion of which are evaluated while others are evaluated by "proactive amateurs"²⁰⁷ – one interviewee for this paper commented that AI was often being held to "unfair and absolute standards, expected to be correct near to 100 per cent of the time". This hinders the adoption of AI solutions which could be 'good enough' for the task at hand.

The Government should outline clear policies on how it expects AI to be evaluated. One measure should be based on whether AI surpasses the current average performance of humans. AI has already surpassed human performance on multiple benchmarks, for example English understanding and visual reasoning.²⁰⁸ Evidence also exists that AI has surpassed human performance in many of the use cases outlined in this report. Octopus Energy's AI chatbots are estimated to do the work of 250 people and has achieved 80 per cent satisfaction rates versus 65 per cent satisfaction rates achieved by humans,²⁰⁹ and AI translation software can reduce the cost of translating a ten page document from £120 to 21p, albeit reduced cost must not equate to reduced quality.²¹⁰

The deterioration of AI models over time is possible, with "concept drift" potentially leading to a situation "where the model may no longer correspond to its new reality".²¹¹ To ensure that when processes are automated they continue to outperform the baseline performance of public servants undertaking that task unaided by AI, public services should continue to operate control groups of cases which are entirely managed by humans.

Recommendation 18: The GDAIS should publish central government guidance on the principles of evaluating AI, including benchmarking the performance of AI against the performance of humans doing the same task, and evaluating against human performance on an ongoing basis once implemented via control groups.

5.5 Trust

Fundamentally, all policies around the development and deployment of AI in the State are designed to ensure public trust is maintained. One interviewee noted that there is a strong perception in government that "early AI technology like police facial recognition... had shifted the public mood" against AI. Yet this is not supported by the data. 32 per cent of respondents in an Office for National Statistics survey were neutral about their feelings towards AI and its impact on society, and a similar proportion of people reported extremely negative (6 per cent) or extremely positive (5 per cent) scores.²¹²

Data may be high quality, guidance may be clear and evaluation may be direct and effective. But if people do not trust AI, it will not be adopted. One way of increasing trust is by ensuring that clear accountability and responsibilities exist with the deployment of AI. One interviewee

²⁰⁷ Patrick King, *An efficiency mindset: prioritising efficiency in Whitehall's everyday work* (Reform, 2023).

²⁰⁸ Stanford Institute for Human-Centered AI, Artificial Intelligence Index Report 2024.

²⁰⁹ Abby Wallace, 'Al Doing the Work of Over 200 People at Octopus, Chief Executive Says', *City A.M.*, 8 May 2023.

²¹⁰ Microsoft, *Harnessing the Power of AI for the Public Sector*.

²¹¹ Daniel Vela et al., 'Temporal Quality Degradation in Al Models', Scientific Reports 12 (2022).

²¹² Office for National Statistics, 'Understanding AI Uptake and Sentiment Among People and Businesses in the UK: June 2023', 16 June 2023.

explained that there was a risk "government AI policy is made in response to the next big crisis, and if that's a crisis in public services, and there is no obvious accountability, then the government will throw the baby out with the bathwater and set AI public services back by many years".

In relation to data, such accountability and responsibility policies already exist. Departments typically have an Information Asset Owner (IAO), a senior individual "responsible for each identified information asset (e.g. database or ICT system) at the appropriate business level within the Department/Agency".²¹³ These individuals are ultimately responsible for the identified information asset within their organisation including compliance with data protection law.²¹⁴

Having a named individual responsible for a given dataset located within a department is the correct place for responsibility and accountability to sit. This individual will have greater knowledge about how that data is being used in practice than anyone else.

But such clear responsibilities and accountability do not currently exist with AI. The government's Ethics, Transparency and Accountability Framework for Automated Decision-Making guidance states that "responsibility and accountability for algorithms and automation, and their associated outcomes should be made clear" with the Secretary of State "ultimately the one accountable for all decision-making in their department".²¹⁵ The Treasury's guidance requires all "business critical" analytical models to be managed by a senior responsible owner (SRO), but this process is not required for models which aren't deemed as "business critical".²¹⁶

Respondents to research conducted by the Alan Turing Institute reported a lack of clarity regarding oversight and significant variation in responsibility exists between departments.²¹⁷ While 24 of the 32 government bodies with deployed AI that responded to an NAO survey always or usually had a named accountable responsible owner for their use cases, fewer than half of these bodies always or usually identified AI use cases at an organisational level before deployment.²¹⁸ Furthermore, while every department now has a Director identified to be responsible for AI, a single individual is unlikely to be able to provide effective accountability for all AI use cases within a department.

The lack of clear responsibilities and accountability makes it more likely that guidance and legislation is not complied with, and future failings in AI set progress back. To simplify this situation a similar model to that applied for data should be adopted, whereby a single individual is responsible – an Algorithm Owner (AlgO) – for each identified algorithm.

This individual should not need to be a specific grade or background. In an innovation-focused area piloting a new idea, an analyst or scientific advisor may be most appropriate for the role whereas when scaling the deployment of an algorithm in a frontline service it may be most appropriate to have an individual directly responsible for leading that service, or a digital/data professional whose role is to support all enterprise systems that service uses.

²¹³ Cabinet Office, 'Government Security', Web Page, 4 August 2023.

²¹⁴ Cabinet Office, 'Guidance on the Information Asset Owner Role', 4 December 2023.

²¹⁵ Cabinet Office, *Ethics, Transparency and Accountability Framework for Automated Decision-Making*, 2023.

²¹⁶ HM Treasury, *Managing Public Money*, 2023.

²¹⁷ Bright et al., 'Generative AI Is Already Widespread in the Public Sector'.

²¹⁸ National Audit Office, Use of Artificial Intelligence in Government.

However, they should need to meet a certain base standard of AI literacy, and complete mandatory training on the requirements of their role.

Recommendation 19: Every central government body should appoint a single named Algorithm Owner, with appropriate Al literacy, for each Al algorithm in use in their organisation and provide them with core training on the requirements of their role.

6. Conclusion

Artificial intelligence is developing far faster than any government can keep up with – particularly a government which is still catching up with the last generation of digital transformation. It will not be a 'silver bullet' for all the challenges which the State faces in the 21st Century. But it has huge promise to improve the stagnant productivity of the public services – not just to do 'more with less', but also to transform the kinds of services available to the public in fundamental ways. More adaptable, responsive and higher-quality public services will all be easier to build with modern technology powered by AI.

However, this transformation will not happen unless government fundamentally rewires itself to adopt the new tools much more quickly. Compared to private industry (particularly the fastest-growing companies) public service adoption of AI is extremely limited, and change is happening too slowly. The public sector is still too risk-averse to even test AI in many cases, and lacks the focus and investment required to see projects through to deployment at scale.

To build momentum, the Government should focus on using AI in areas where there is a strong evidence base, a quick route to adoption at scale, and which add value in the most challenging parts of the State. Proving the government can successfully use AI, and reap widespread benefits from doing so, is essential to making the case for further adoption.

In practice, this means reorienting the Government's approach – greater central leadership to drive AI adoption throughout public services; funding which can be flexibly deployed, quickly, to sustain and scale up successful pilots; more in-house capability; and procuring much more flexibly to reduce the reliance on a small number of suppliers.

Crucially, government needs to be willing to take smart risks in deploying AI. This is particularly true in the early stages of AI development, when the current risk tolerance is disproportionate to the actual level of risk. Stringent evaluation is essential before AI is deployed in enterprise-level services, but the required performance standards should not be unrealistic – in most cases, AI software which performs as well as humans would be more than adequate.

Government must assess use cases against the risks of continuing with the status quo – failing public services, declining accessibility, rising costs, and in some cases high levels of bias.

This new approach requires accepting that transformed public services will not happen organically, and innovation needs to be led by the centre in partnership with frontline services. And it requires establishing parity between the risk of adopting AI with the risks of not doing so.

Appendix 1: Defining AI

For the purposes of this paper, *Reform* use the definition of AI used by the Defence Science and Technology Laboratory (Dstl):

"Theories and techniques developed to allow computer systems to perform tasks normally requiring human or biological intelligence".²¹⁹

This is different to, but overlaps with, fields such as Machine Learning and Data Science, with the relationships shown in Figure 8.

Figure 8: Dstl definition of the relationship between Artificial Intelligence, Machine Learning and Data Science



Source: Artificial Intelligence, Data Science and (mostly) Machine Learning, The Dstl Biscuit Book, 1st edition revised v1_2

Within the field of AI, it can be useful to distinguish between the concepts of 'traditional' or 'narrow' AI, and Generative AI.

"Traditional, or narrow, AI is primarily involved in analysing and classifying the information it's trained on. It excels in tasks involving processing input data and sorting this data into pre-designated categories, for example, determining whether an email is spam or not based on its content.

Generative AI aims to create new data that mimics the input it's been trained on, such as large language models (LLMs) generating original text in a particular style that was previously only possible with human creativity."²²⁰

²¹⁹ Defence, Science and Technology Laboratory, *The Dstl Biscuit Book: Artificial Intelligence, Data Science and (Mostly) Machine Learning.*

²²⁰ London Office of Technology and Innovation & Faculty, *Opportunities for AI in Housing Services*, 2024.

Bibliography

- Ada Lovelace Institute. *AI Assurance?* Assessing and Mitigating Risks Across the AI Lifecycle, 2023. AI Safety Institute. 'Careers'. Web Page, n.d.
- Alan Turing Institute. Generative AI Is Already Widespread in the Public Sector, 2024.
- Baumol, William J., and William G. Bowen. *Performing Arts: The Economic Dilemma*. New York: The Twentieth Century Fund, 1966.
- Beam. 'Transform Social Workers' Productivity with Al'. Beam, n.d.
- Belfield, Haydn. Great British Cloud and BritGPT: The UK's AI Industrial Strategy Must Play to Our Ambitions. Labour for the Long Term, 2023.
- Bennett, Andrew, and Chris Yiu. *Transforming Government for the 21st Century*. Tony Blair Institute for Global Change, 2019.
- Blackall, Molly. 'One Asylum Appeal Lodged Every Hour on Average in Britain'. *The i*, 18 February 2024.
- Bright, Jonathan, Florence Enock, Saba Esnaashari, John Francis, Youmna Hashem, and Deborah Morgan. 'Generative AI Is Already Widespread in the Public Sector'. *arXiv Preprint:* 2401.01291, 3 January 2024.
- British Medical Association. 'NHS Diagnostics Data Analysis'. Web Page, 4 October 2023.
- Browne Jacobson. 'Procurement Bill: Competitive Flexible Procedure, How Will This Work in Practice?', 27 September 2023.
- Bryom, Natalie, Rachel Coldicutt, and Sarah Gold. *People First, Always: Delivering Better, Cheaper, More Accessible Public Services*, 2024.
- Burgess, Matt. 'London Underground Is Testing Real-Time AI Surveillance Tools to Spot Crime'. *WIRED*, 8 February 2024.
- Cabinet Office. 'Civil Service Fast Stream Exceeds Target to Boost STEM Expertise Across Government'. Web Page, 25 October 2023.
- ———. Ethics, Transparency and Accountability Framework for Automated Decision-Making, 2023.
- . 'Government Professions'. Web Page, 16 February 2024.
- . 'Government Security'. Web Page, 4 August 2023.
- -------. 'Guidance on the Information Asset Owner Role', 4 December 2023.
- ———. 'Whitehall Set to Bring in AI and Data Experts under Plans to Turbocharge Productivity', 19 July 2023.

Caldicott, Fiona. 'Data Sharing During This Public Health Emergency'. Web Page, 3 April 2020. Canadian Digital Service. 'Digital Products for the Government of Canada'. Web Page, n.d.

- ———. Guidance on the Legacy IT Risk Assessment Framework, 2023.
- ——. Managing Legacy Technology, 2019.
- ———. Transforming for a Digital Future: 2022 to 2025 Roadmap for Digital and Data Updated September 2023, 2023.
- ———. Transforming for a Digitial Future: 2022 to 2025 Roadmap for Digital and Data Original, 2022.
- Cespedes, Frank, and Jacco van der Kooij. 'The Rebirth of Software as a Service'. *Havard Business Review*, 18 April 2023.
- Civic Future. 'Should the State Play More of a Role in the Economy?' Web Page, 29 June 2024. Civil Service. 'The Civil Service Code'. Webpage, 16 March 2015.
- 'Civil Service People Plan 2024-2027', 10 January 2024.
 - https://assets.publishing.service.gov.uk/media/65eaf2335b6524cb5ff21aa8/Civil_Service_People_Plan_2024-2027_Web_FV.pdf.
- Constanz, Jo. 'OpenAl Engineers Earning \$800,000 a Year Turn Rare Skillset Into Leverage'. Yahoo *Finance*, 22 November 2023.
- Copeland, Eddie. 'It's Time We Had a Local Authority Sandbox'. LOTI (blog), 13 January 2024.

Crown Commercial Service. Management Consultancy Framework Three (MCF3), 2021.

Dalrymple, David 'davidad'. Safeguarded AI: Constructing Guaranteed Safety. ARIA, 2024.

Davies, Matt, and Elliot Jones. Foundation Models in the Public Sector. Ada Lovelace Institute, 2023.

Central Digital and Data Office. 'Digital and Technology Spend Controls (Version 5)'. Web Page, 21 July 2023.

Defence, Science and Technology Laboratory. *The Dstl Biscuit Book: Artificial Intelligence, Data Science and (Mostly) Machine Learning*, 2019.

Dent, Anna. Automating Public Services: A Careful Approach. Promising Trouble, 2024.

Department for Environment Food & Rural Affairs. Environmental Principles Policy Statement, 2023.

Department for Science, Innovation and Technology. A Pro-Innovation Approach to AI Regulation, 2023.

———. 'AI Expert to Lead Action Plan to Ensure UK Reaps the Benefits of Artificial Intelligence'. Web Page, 26 July 2024.

———. 'DSIT Bolstered to Better Serve the British Public Through Science and Technology'. Web Page, 8 July 2024.

------. 'Global Leaders Agree to Launch First International Network of AI Safety Institutes to Boost Cooperation of AI', 21 May 2024.

——. National AI Strategy - AI Action Plan, 2022.

Department of Health and Social Care. 'Daily Tests Processed and Testing Capaity (UK): 20 March to 22 September 2020'. Web Page, 24 September 2020.

Dohmke, Thomas. 'The Economic Impact of the AI-Powered Developer Lifecycle and Lessons From GitHub CoPilot'. Web Page, 14 May 2024.

Driver and Vehicle Licensing Agency. 'DLVA's IT Transformation – Our Story so Far'. Web Page, 3 August 2022.

Dunton, Jim. 'Cabinet Office Flags Digital Successes with Three-Year Roadmap'. *Civil Service World*, 29 September 2023.

DWP Digital. 'Intelligent Automation Garage: Using Automation for Good'. Web Page, n.d.

European Parliament and The Council of the European Union. *General Data Protection Regulation*, 2016.

Faculty, London Office of Technology and Innovation. *Opportunities for AI in Adult Social Care Services*, 2024.

—. Opportunities for AI in Housing Services, 2024.

Ferber, Alona. 'Louise Casey: The Tories Are Done - and They Need to Go Away for a While'. *New Statesman*, 29 November 2023.

Financial Conduct Authority. Implementing Technology Change, 2021.

Fisher, Lucy. 'UK Government to Trial "red Box" AI Tools to Improve Ministerial Efficiency'. *Financial Times*, 28 February 2024.

Freeguard, Gavin, and Paul Shepley. *Legislation to Support Data Sharing*. Institute for Government, 2022.

Gartside, Ben. 'ChatGPT: Civil Servants Warned Not to Use AI Chatbots to Write Policies and Carry out Government Work'. *iNews*, 21 February 2023.

Gilbert, Laura. Changing Government One Datastream at a Time. Heywood Quarterly, 2024.

Government Digital Service. 'Buying and Selling on the Digital Marketplace', November 2020.

------. 'How to Pay for Digital Outcomes and Specialists Services', July 2016.

——. 'Service Manual', n.d.

------. 'Service Standard', n.d.

Greenway, Andrew, and Tom Loosemore. The Radical How. Public Digital, Nesta, 2024.

Harwich, Eleonora, and Kate Laylock. *Thinking on Its Own: AI in the NHS*. Reform, 2018.

Hellen, Nicholas. "Robowardens" to Scan Whole Streets for Parking Violations in Minutes'. *The Sunday Times*, 27 July 2024.

Hill, Joe, Charlotte Pickles, and Sean Eke. *Making the Grade: Prioritising Performance in Whitehall*. Reform, 2024.

HM Government. CONTEST: The United Kingdom's Strategy for Countering Terrorism, 2018. ———. Digital Economy Act 2017 (2017).

——. National AI Strategy, 2021.

HM Government, and Central Digital and Data Office. *Generative AI Framework for HM Government*, 2024.

HM Land Registry. 'HM Land Registry: Processing Times', 19 July 2024.

HM Treasury. '£1.8 Billion Benefits through Public Sector Productivity Drive', 2 March 2024.

------. 'Guide to Developing the Project Business Case', 2018.

——. Managing Public Money, 2023.

Howes, Sally, and Tess Kidney Bishop. 'The Hidden Obstacles to Government Digital Transformation', October 2018.

- Information Commissioner's Office. 'ICO Issues Reprimand to the Metropolitan Police Service for Inadequate Handling of Files Related to Organised Crime Groups'. Web Page, 16 March 2023.
- Institute for Government. Why Is Civil Service Reform so Hard? Sir John Kingman in Conversation with Bronwen Maddox, 2020.

losad, Alexander, David Railton, and Tom Westgarth. *Governing in the Age of AI: A New Model to Transform the State*. Tony Blair Institute for Global Change, 2024.

- Jeffrey, Olivia, Jessamy Bloom, Timothy Eales, Nick Morgan, Robert Patman, William Gould, Rajeevan Balachandran, and Peppa Pancheva. *Evaluation Report on Grip and Bespoke-Funded Hot Spot Policing*. Home Office, 2024.
- Joint Research Centre. Al Watch: European Landscapes on the Use of Artificial Intelligence by the Public Sector, 2022.
- Jung, Carsten, and Bhargav Srinivasa Desikan. *Transformed by AI*. Institute for Public Policy Research, 2024.
- Kelion, Leo. 'Excel: Why Using Microsoft's Tool Caused Covid-19 Results to Be Lost'. *BBC News*, 5 October 2020.
- Kelly, Jack. 'Klarna's Al Assistant Is Doing The Job Of 700 Workers, Company Says'. *Forbes*, 4 March 2024.
- Kendall, Michelle, Daphne Tsallis, Chris Wymant, Andrea Di Francia, Yakubu Balogun, Xavier Didelot, Luca Ferretti, and Christophe Fraser. 'Epidemiological Impacts of the NHS COVID-19 App in England and Wales Throughout Its First Year'. *Nature Communications* 14 (2023).
- King, Patrick. An Efficiency Mindset: Prioritising Efficiency in Whitehall's Everyday Work. Reform, 2023.
- Lane Fox, Martha. Directgov 2010 and Beyond: Revolution Not Evolution, 2010.
- Lofgren, Eric. 'Explaining the Valley of Death in Defense Technology'. *Acquisition Talk* (blog), 9 December 2019.
- Madsen, Eric. 'Farmer's Diagram, or F-N Curve. Representing Society's Degree of Catastrophe Aversion'. 24 July 2022. Risk Engineering, n.d.
- Maltby, Paul. 'Reasons to Be Optimistic About DSIT and the New Digitial Centre of Government'. Web Page, 24 July 2024.
- Markson, Tevye. 'Scale of Government Legacy IT Systems at "Critical" Risk Revealed'. *Civil Service World*, 10 January 2024.
- Maude, Francis. Independent Review of Governance and Accountability in the Civil Service, 2023.
- McKeon, Christopher. 'Let 1,000 AI Pilots Flourish to Help Public Sector, Says Labour MP'. Independent, 9 August 2023.
- McNeil, Rupert. Oral Evidence: Civil Service Human Resources, HC 1399. Public Administration and Constitutional Affairs Committee, 2023.
- McRae, Isabella. 'DWP Ramps up AI Use to "Bring the Future to the Welfare System" and Push People Into Work'. *The Big Issue*, 21 May 2024.
- Memon, Amina, Zoe Given-Wilson, Derya Ozkul, Karen McGregor Richmond, Julia Muraszkiewicz, Ella Weldon, and Cornelius Katona. 'Artificial Intelligence (AI) in the Asylum System'. *Medical, Science and the Law* 64, no. 2 (2024).
- Meta. 'Introducing Llama 3.1: Our Most Capable Models to Date'. Meta AI (blog), 23 July 2024.
- Microsoft. Harnessing the Power of AI for the Public Sector, 2024.
- Microsoft, and Public First. Unlocking the UK's AI Potential, 2024.

Ministry of Justice. 'Tribunal Statistics Quarterly: October to December 2023'. Web Page, 14 March 2024.

- Murray, Dale. 'Open Source and Security: Why Transparency Now Equals Strength'. *Network Security* 2020, no. 7 (n.d.).
- National Audit Office. Challenges in Using Data Across Government, 2019.
- ——. Digital Transformation in Government: Addressing the Barriers to Efficiency, 2023.
- ———. The Challenges in Implementing Digital Change, 2021.
- ———. 'The National Law Enforcement Data Programme'. Web Page, 10 September 2021. ———. Use of Artificial Intelligence in Government, 2024.
- NHS England. 'Using an Al-Driven Dictation Platform to Free up Clinicians' Time'. Web Page. n.d.
- Nilsson, Patricia. 'First UK Police Force to Try Predictive Policing Ends Contract'. *Financial Times*, 26 November 2018.
- OCR. 'Examiners'. Web Page, n.d.
- Office for Budget Responsibility. Economic and Fiscal Outlook March 2024, 2024.

- Office for National Statistics. 'Public Service Productivity, Quarterly, UK: January to March 2024'. Web Page, 15 July 2024.
 - —. 'UK Government Debt and Deficit: December 2023'. Web Page, 30 April 2024.

—. 'Understanding AI Uptake and Sentiment Among People and Businesses in the UK: June 2023', 16 June 2023.

Open Government Products. 'About Us'. Web Page, n.d.

- OpenAI. 'GPT4o- Demo'. Web Page, May 2024.
- O'Reilly, Tim. 'The UK Needs to Double Down on the GDS, Not Dismantle It'. Web Page, 16 August 2016.
- Pickles, Charlotte, and James Sweetland. *Breaking Down the Barriers: Why Whitehall Is so Hard to Reform*. Reform, 2023.

Public Accounts Committee. Digital Services at the Border, 2021.

- ——. Digital Transformation in Government: Addressing the Barriers to Efficiency, 2023.
- ------. The Dismantled National Programme for IT in the NHS, 2013.
- PUBLIC, Paradigm Junction. Buying Generative AI in Government, 2024.
- Ram, Aliya, and Sarah Neville. 'High-Profile Health App Under Scrutiny After Doctors' Complaints'. *Financial Times*, 13 July 2018.
- Sasse, Tom, Rhys Clyne, and Sachin Savur. 'Asylum Backlog'. Web Page, 24 February 2023.
- Smith, Mark. 'Can We Predict When and Where a Crime Will Take Place?' *BBC News*, 30 October 2018.
- Sorgi, Gregorio. 'EU Gives More Power to Al Translation Machines'. Politicio, 15 June 2023.
- Spahr, Thomas W. *Raven Sentry: Employing AI for Indications and Warnings in Afghanistan*. Vol. Volume 54. Parameters. The US Army War College Quarterly, 2024.
- Stanford Institute for Human-Centered AI. Artificial Intelligence Index Report 2024, 2024.
- TaxPayers' Alliance. NHS Spending on Interpreters and Translation Services, 2023.
- The Behavioural Insights Team. A Review of Optimism Bias, Planning Fallacy, Sunk Cost Bias and Groupthink in Project Delivery and Organisational Decision Making. Department for Transport, 2017.
- The NHS AI Lab. 'AI Diagnostic Fund'. Web Page, n.d.
- ------. 'Using AI to Improve Back Office Efficiency in the NHS'. Web Page, 11 February 2022.
- The White House. 'Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence'. Web Page, 30 October 2023.
- Thomas, Richard, and Mark Walport. Data Sharing Review Report, 2008.
- Trendall, Sam. 'Al Could Save £5bn a Year and Take Workload from "Tens of Thousands" of Jobs, Government Ops Chief Says'. *Public Technology* (blog), November 2023.
- Turner, Cerys. 'GCSEs 2024: Exam Board to Trial AI in Summer Exams'. *TES Magazine*, 15 March 2024.
- Vela, Daniel, Andrew Sharp, Richard Zhang, Trang Nguyen, An Hoang, and Oleg Pianykh. 'Temporal Quality Degradation in Al Models'. *Scientific Reports* 12 (2022).
- Walden, Chris. 'How Individual Contributors Can Become Brilliant Technical Leaders', 24 June 2020.
- Wallace, Abby. 'Al Doing the Work of Over 200 People at Octopus, Chief Executive Says'. *City A.M.*, 8 May 2023.
- Weiss, Andrew. 'Empowering Developers in Government with GitHub Copilot'. Github, n.d.
- Willetts, David. Independent Review of the DSIT Business Case and Approvals Process, 2024.

Woodhouse, John. Data Protection and Digital Information (No.2) Bill: Progress of the Bill. House of Commons Library, 2023.

Yeung, Karen. 'Written Evidence to the Public Accounts Committee: Use of Artificial Intelligence in Government'. Web Page, May 2024.





@reformthinktank

info@reform.uk

www.reform.uk