Foundation models in the public sector

Foundation models are a form of artificial intelligence (AI) designed to produce a wide variety of outputs that are being adopted across various sectors. They are capable of a range of tasks and applications, such as text, image or audio generation.[1] Notable examples are OpenAI's GPT-3 and GPT-4, which underpin the conversational tool ChatGPT.

As AI technologies advance rapidly, government must consider carefully how to use foundation models in the public sector responsibly and beneficially. This paper provides policymakers and public-sector leaders with information to help them to do this.

---

1   For more details on foundation models, read the Ada Lovelace Institute's explainer 'What is a foundation model?' https://www.adalovelaceinstitute.org/resource/foundation-models-explainer/

Ada Lovelace Institute

We start with an overview of foundation modelsi and their potential use cases in central and local government in the UK. We then consider their risks and opportunities in the public sector, as highlighted by public-sector leaders, researchers and civil society.

Finally, we explore the principles, regulations and practices, such as impact assessments, monitoring and public involvement, necessary to deploy foundation models in the public sector safely, ethically and equitably.

Executive summary

Foundation models are already being integrated into commonly used applications: Google and Microsoft's Bing embed them in search engines, Adobe's Photoshop integrates image generation models,[2] and companies like Morgan Stanley use large language models (LLMs) for internal knowledge search and retrieval.[3]

There is some optimism in policy, public sector and industry settings about the potential for these models to enhance public services by:[4]

- automating review of complex contracts and case files (document analysis)
- catching errors and biases in policy drafts (decision support)
- powering real-time chatbots for public enquiries (improvements in management of public enquiries)
- consolidating knowledge spread across databases into memos (knowledge management).

However, there are also risks around issues like biases, privacy breaches, misinformation, security threats, overreliance, workforce harms and unequal access.

To produce this paper, we completed a desk-based review of the academic literature, news reports, transparency documents and similar sources outlining current or near-term government use of foundation models. We complemented this with two expert roundtables bringing together government product teams, industry, civil society and academia to provide input on the research questions below. We commissioned the AI Law Consultancy to conduct a desk-based review of UK legislation, regulations and guidance, identifying where foundation models and related concepts were directly addressed or mentioned. This research report is included in Appendix 2.

---

2   'Generative AI for Creatives - Adobe Firefly' https://www.adobe.com/uk/sensei/generative-ai/firefly.html accessed 15 August 2023.

3   'Key Milestone in Innovation Journey with OpenAI' (Morgan Stanley) https://www.morganstanley.com/press-releases/key-milestone-in-innovation-journey-with-openai accessed 15 August 2023.

4   For more detail on potential use cases, please see Appendix 1.

Key findings

- Foundation models rely on large-scale data and compute for training. Their capabilities centre on text, image and data analysis or data generation. Prominent examples include chatbots like ChatGPT or Claude, and image generators like Midjourney.

- Potential uses include document analysis, decision support, policy drafting and public knowledge access, according to those working in or with the public sector.

- Public services should carefully consider the counterfactuals to implementing foundation models. This means comparing proposed use cases with more mature and tested alternatives that might be more effective or provide better value for money.

- Evaluating these alternatives should be guided by the principles of public life.

- Risks include biases, privacy breaches, misinformation, security threats, overreliance, workforce harms and unequal access.

- It is vital to mitigate these risks through monitoring, internal and independent oversight, and engaging with those affected by the technologies.

- Existing guidance and impact assessments provide baseline governance for using foundation models but may need enhancement. Small pilots, independent auditing and public involvement can also minimise risks.

- Government should invest in skills and address technical dependencies.

- Government could consider options like funding domestic data centres and updates to procurement guidelines for AI systems.

- As foundation models' capabilities evolve and market dynamics change, there will be new opportunities for public-interest-driven innovation, but new risks also need to be anticipated to ensure effective governance.

Research questions

To support policymakers' decision-making and public understanding around foundation models in the public sector, we answer these questions:

1.	What is a foundation model, and how do we define the different terms around this technology?
2.	How are foundation models being deployed within government, formally and informally? What further uses are in development or are being considered?
3.	What problems or opportunities in public service delivery do local and central government see foundation models solving? Are there better, more robust and better regulated tools already available?
4.	What are the risks, unintended consequences and limitations of foundation models for public services and their autonomy?
5.	How should government use of these technologies be governed? What hard rules or guidance should governments follow?

Introduction

The rapid evolution of artificial intelligence (AI) has brought us to a point where foundation models are being adopted across various sectors. Foundation models are a form of AI designed to produce a wide and general variety of outputs, capable of a range of tasks and applications, such as text, image or audio generation.[5] Notable example are OpenAI's GPT-3 and GPT-4, foundation models that underpin the conversational tool ChatGPT.

Following the launch of large language model (LLM) interfaces like ChatGPT and image generators like DALL-E 2 and Stable Diffusion, foundation models are more widely accessible than ever. Google and Microsoft's Bing are embedding the models into everyday search, tools like Photoshop are integrating image generation models, and companies like Morgan Stanley use LLMs for internal knowledge search and retrieval.

Glossary

Term	Meaning	Origin/context/notes
Foundation model (see also 'GPAI')	Described by researchers at Stanford University Human-Centered Artificial Intelligence as: 'AI neural network trained on broad data at scale that can be adapted to a wide range of tasks' [6] [7]	Coined by Stanford University Human-Centered Artificial Intelligence. EU AI Act: Foundation models serve as a base model for other AI systems that will be 'fine-tuned' from it. They function as platforms for a wave of AI applications, including generative AI.[8] This term is often used interchangeably with GPAI.

---

5   Ada Lovelace Institute (n 1).

6   Reflections on Foundation Models' (Stanford Institute for Human-Centered Artificial Intelligence) https://hai.stanford.edu/news/reflections-foundation-models

7   'Texts Adopted – Artificial Intelligence Act - Wednesday, 14 June 2023' https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html accessed 23 November 2023

8   Responses to NTIA's Request for Comment on AI Accountability Policy' (Stanford Institute for Human-Centered Artificial Intelligence) https://hai.stanford.edu/responses-ntias-request-comment-ai-accountability-policy

GPAI (see also 'foundation model')	'General purpose AI system' means an AI system that can be used in and adapted to a wide range of applications for which it was not intentionally and specifically designed'[9]	EU AI Act Under the EU AI Act, the term GPAI refers to an AI system which can be adapted to a wide range of applications. This term is often used interchangeably with foundation model.
Generative AI	A type of AI system that can create a wide variety of data, such as images, videos, audio, text and 3D models[10]	
Large language model (LLM)	Large language models are type of AI system trained on massive amounts of text data that can generate natural language responses to a wide range of inputs.	Increasingly, these large models are multimodal. For example, while GPT-4 is primarily text-based and only gives text-based outputs, it can use both text and images simultaneously as an input.

Many public- and private-sector organisations view foundation models with optimism. Some within government product teams and suppliers of data and machine learning services to government see foundation models as a way to tackle some of the big problems in society.

Central government departments, local authorities and other public-sector organisations are already considering using these systems in various ways, such as in decision-making, disseminating information and research, enabling wider access to data, delivering services and monitoring service provision, and are discussing this with private-sector suppliers such as Faculty and Palantir.

They believe foundation models can help to tackle the cost-of-living crisis, drive growth through innovation, solve the complexity of data sharing across government and make local government services more efficient.

However, there are risks around issues like bias, privacy, security, environmental impact and workforce displacement. Government needs to consider carefully how these technologies can be deployed responsibly and beneficially. This paper aims to help policymakers and public-sector leaders to do this.

To inform the report, the Ada Lovelace Institute undertook a rapid review of foundation models in the UK public sector to identify current and potential uses and what governance and oversight should look like. We also reviewed academic and 'grey' (policy and industry) literature and convened

---

9   https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/CJ40/DV/2023/05-11/ConsolidatedCA_IMCOLIBE_AI_ACT_EN.pdf

10 https://generativeai.net/

two roundtables, with government product teams and with representatives from civil society and industry. We also commissioned the AI Law Consultancy to analyse relevant legislation, regulations and guidelines on AI, and we drew on wider conversations with government officials and on relevant roundtables hosted by other organisations.

The report defines foundation models, the social context of their deployment and their potential applications in the public sector. It considers risks highlighted by public-sector leaders, researchers and civil society. It then explores relevant current principles, regulation and laws, before looking at additional oversight mechanisms needed to ensure that the government deploys foundation models in the public sector safely, ethically and equitably.

A note on terminology

We initially focused on 'general-purpose AI' systems, following terminology used in the European Parliament's draft EU AI Act. But it was soon clear that terminology in this area is inconsistent and still evolving. Terms used include:

- general-purpose AI (GPAI)
- foundation model
- generative AI
- frontier model
- artificial general intelligence (AGI)
- large language model (LLM)

The public, policymakers, industry and the media need a shared, up-to-date understanding of the terminology, to support clearer understanding and make communication and decision-making more effective. We have developed an explainer to help provide this common understanding.[11]

Throughout the report we primarily use the term 'foundation model', coined at Stanford University and adopted by the UK Government in its policy outputs and in the mission statement of the Government's Frontier AI Taskforce (formerly the Foundation Model Taskforce).[12]

---

11   Ada Lovelace Institute (n 1).

12   A Stanford University report defined foundation models as 'any model that is trained on broad data (generally using self-supervision at scale) that can be adapted [...] to a wide range of downstream tasks'. See: Rishi Bommasani and others, 'On the Opportunities and Risks of Foundation Models' (arXiv, 12 July 2022) 3 http://arxiv.org/abs/2108.07258 accessed 30 January 2023."plainCitation":"Rishi Bommasani and others, 'On the Opportunities and Risks of Foundation Models' (arXiv, 12 July 2022

How to read this review

...if you are part of a product team in central or local government:

- The **'Current use and future application of foundation models in government and public services'** section includes common challenges. The '**Risks of foundation models in the public sector'** section summarises risks to consider when procuring or deploying foundation models. The '**Existing legislation, regulation and guidance'** section covers existing tools like impact assessments that can be used to mitigate risks in deploying foundation models. The **'What else can be done to govern foundation models in the UK public sector?'** section suggests further steps you could take like mandating audits and public involvement to ensure responsible oversight.

...if you are a regulator overseeing public sector use of foundation models:

- The **'Risks of foundation models in the public sector'** section will help identify where regulatory powers may need strengthening. The '**Existing legislation, regulation and guidance'** section summarises relevant legislation and guidelines, providing an overview of the regulatory landscape. **'What else can be done to govern foundation models in the UK public sector?'** contains potential changes to the governance of public-sector foundation models, including a proactive review of policies and expansion of regulators' powers.

...if you work for a technology company that sells into the public sector:

- The sections on **'Challenges in deploying foundation models in the public sector'** and the '**Risks of foundation models in the public sector'** outline practical and ethical concerns around the use of foundation models. Considering these can help you develop products suited to public sector needs.
- The **'Procurement and technical lock-in'** subsection suggests how procurement might need to set social responsibility standards for your products.

... if you are from a civil society organisation:

- The **'Risks of foundation models in the public sector'** section covers potential harms to vulnerable groups that civil society could help identify and mitigate.
- The **'What else can be done to govern foundation models in the UK public sector?'** section suggests changes to foundation model governance, including public participation.

Background

This section provides an overview of foundation models, current discourse around them, their growing use in the UK public sector and their potential applications.

What is a foundation model?

Foundation models are a form of AI designed for a wide range of possible applications and ability to achieve a range of distinct tasks, including translating and summarising text, generating a first draft of a report from notes, or responding to a query from a member of the public.

These applications often use the models without substantial modification. However, the model may also be fine-tuned. Fine-tuning is the process of additional context-specific training. For example, OpenAI fine-tuning the GPT-3.5 and GPT-4 families of models used in ChatGPT.[13]

Foundation models can take in one or many different types of inputs, such as text, images, videos or sound, and respond with one or many different types of outputs. A foundation model could generate an image based on text input, or a video based on an image and text combined.

They can be directly available to consumers in standalone systems, as are GPT-3.5 or GPT-4 through the ChatGPT interface. Or they can be the 'building block' of hundreds of single-purpose AI systems.

Current foundation models are defined by their scale. They are trained on billions of words of text, use millions of pounds worth of compute per training run,[14] and rely on transfer learning (applying knowledge from one task to another). Future foundation models may not necessarily have these properties.[15] In contrast, narrow AI applications are those trained for a specific task and context, making it difficult for them to be reused for new contexts.

Since ChatGPT was released in November 2022, foundation models and their applications have attracted substantial media, consumer and investor attention. ChatGPT gained 100 million

13   Sabrina Küspert, Nicolas Moës and Connor Dunlop, 'The Value Chain of General-Purpose AI' (Ada Lovelace Institute, 10 February 2023) https://www.adalovelaceinstitute.org/blog/value-chain-general-purpose-ai/>accessed 27 March 2023.

14   A training run refers to a critical production process for general purpose AI models that require computing resources.

15   Risto Uuk, 'General Purpose AI and the AI Act' (Future of Life Institute 2022) https://artificialintelligenceact.eu/wp-content/uploads/2022/05/General-Purpose-AI-and-the-AI-Act.pdf accessed 26 March 2023.

monthly active users in two months, making it the fastest-growing consumer application in history.[16] Competitors have collectively raised billions of dollars in investment to develop foundation models.[17]

How are foundation models developed, tested and deployed?

Developers and deployers of AI systems have various distinct responsibilities, from defining the problem, to collecting, labelling and cleaning data, to model training and 'fine-tuning', to testing and deployment (see Figure 1 below). Some developers undertake the entire process in-house, but these different activities are often carried out by multiple actors in a supply chain.[18]

Every AI system has a different supply chain, depending on sector, use case, whether the system is developed in-house or procured, and how it is made available to users, as we discuss below.

---

16   Krystal Hu, 'ChatGPT Sets Record for Fastest-Growing User Base - Analyst Note' Reuters (2 February 2023) https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/ accessed 23 August 2023.

17   Krystal Hu and Manya Saini, 'AI Startup Cohere Raises Funds from Nvidia, Valued at $2.2 Billion' Reuters (8 June 2023) https://www.reuters.com/technology/ai-startup-cohere-raises-270-mln-nvidia-backed-funding-round-2023-06-08/ accessed 23 August 2023; Krystal Hu, Jaiveer Shekhawat and Krystal Hu, 'Google-Backed Anthropic Raises $450 Mln in Latest AI Funding' Reuters (23 May 2023) https://www.reuters.com/markets/deals/alphabet-backed-ai-startup-anthropic-raises-450-million-funding-freeze-thaws-2023-05-23/ accessed 23 August 2023; Niket Nishant and Krystal Hu, 'Microsoft-Backed AI Startup Inflection Raises $1.3 Billion from Nvidia and Others' Reuters (29 June 2023) https://www.reuters.com/technology/inflection-ai-raises-13-bln-funding-microsoft-others-2023-06-29/ accessed 23 August 2023; Joyce Lee, 'Google-Backed Anthropic Raises $100 Mln from South Korea's SK Telecom' Reuters (14 August 2023) https://www.reuters.com/technology/google-backed-anthropic-raises-100-mln-south-koreas-sk-telecom-2023-08-14/ accessed 23 August 2023.

18   Alex Engler and Andrea Renda, 'Reconciling the AI Value Chain with the EU's Artificial Intelligence Act' (Centre for European Policy Studies 2022) 2–3 https://www.ceps.eu/ceps-publications/reconciling-the-ai-value-chain-with-the-eus-artificial-intelligence-act/

Figure 1: The AI model lifecycle
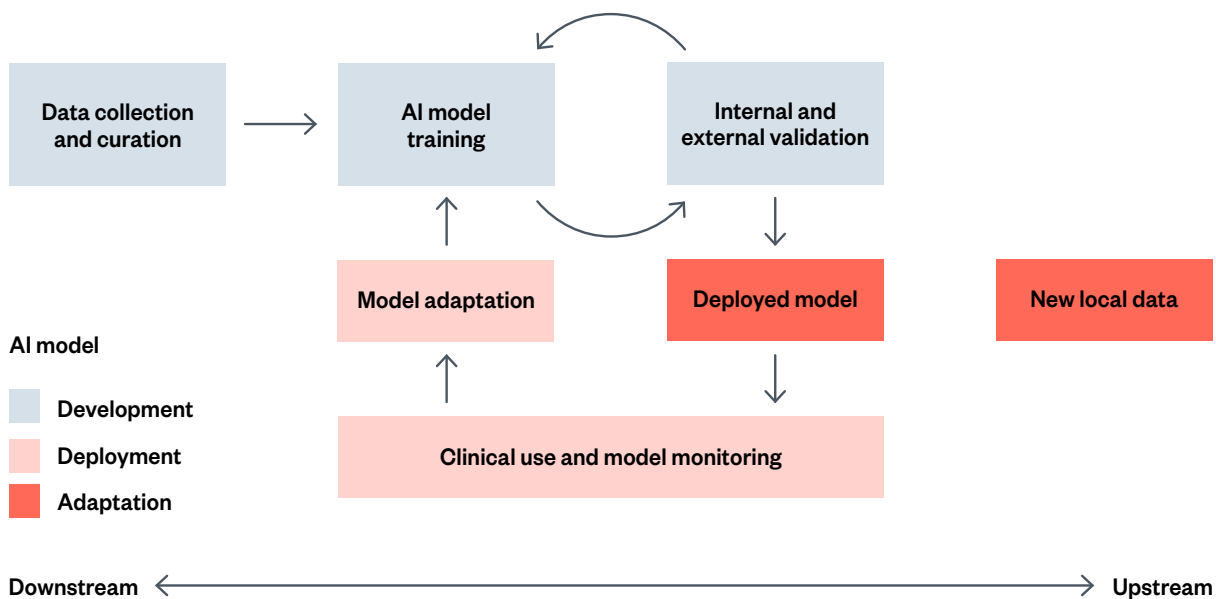
How might foundation models be made accessible to government?

To date, foundation models have mostly been hosted on cloud computing platforms, such as Amazon Web Services, Microsoft Azure and Google Cloud, and made accessible via an application programming interface (API) to other developers, who can fine-tune them using their own data. Many end users access such systems via existing tools, such as operating systems, browsers, voice assistants and productivity software (e.g. Microsoft Office, Google Workspace).

Many businesses build context-specific AI applications on top of a foundation model API. For example, private-sector companies have started to build products underpinned by the OpenAI's GPT-4 API, including:

- Microsoft's BingChat for internet search and retrieval[19]
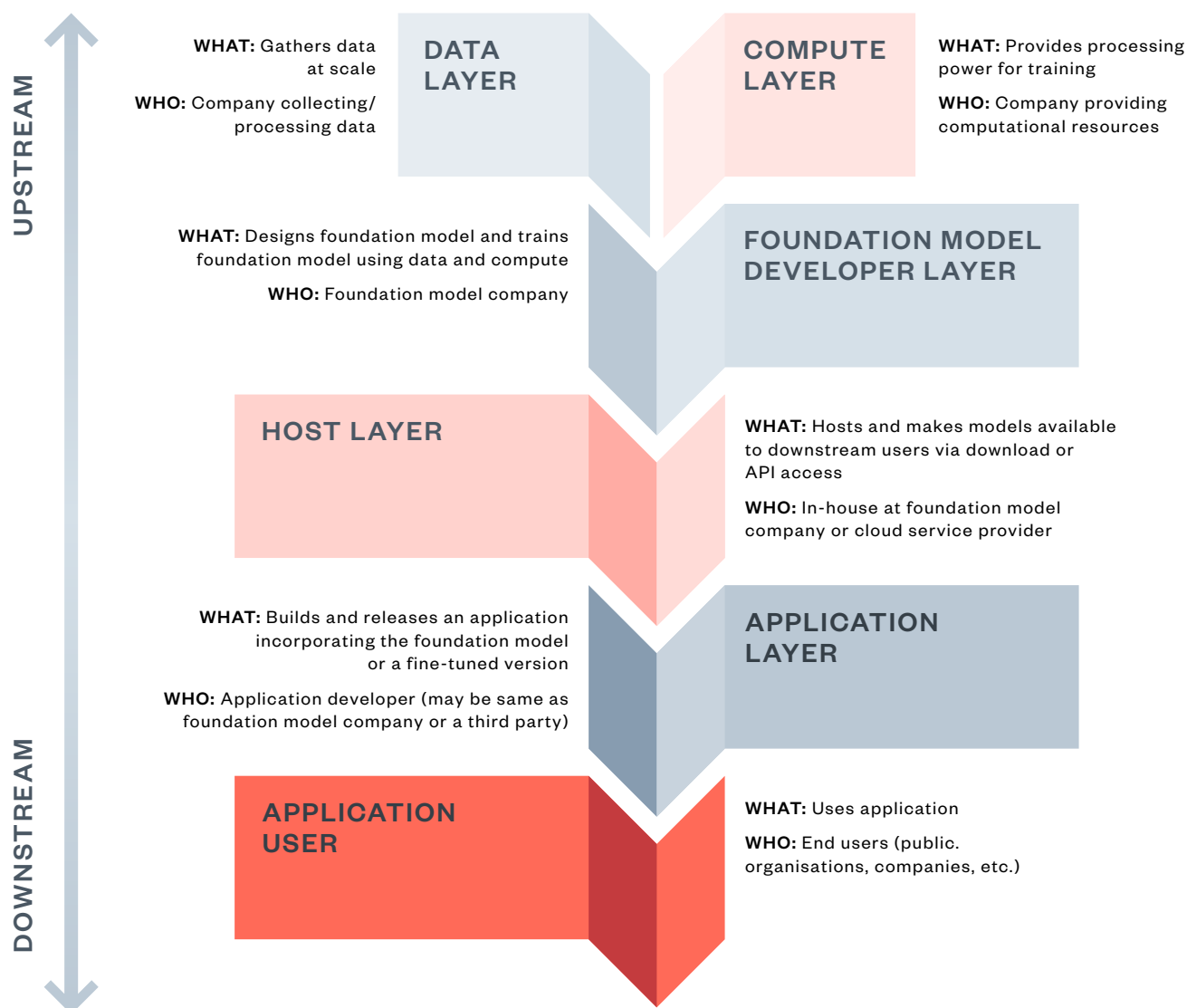- Duolingo Max[20] and Khan Academy's Khanmigo[21] for personalised education

---

19   Yusuf Mehdi, 'Reinventing Search with a New AI-Powered Microsoft Bing and Edge, Your Copilot for the Web' (The Official Microsoft Blog, 7 February 2023) https://blogs.microsoft.com/blog/2023/02/07/reinventing-search-with-a-new-ai-powered-microsoft-bing-and-edge-your-copilot-for-the-web/ accessed 26 March 2023.

20   Duolingo Team, 'Introducing Duolingo Max, a Learning Experience Powered by GPT-4' (Duolingo Blog, 14 March 2023) https://blog.duolingo.com/duolingo-max/ accessed 26 March 2023.

21   Sal Khan, 'Harnessing GPT-4 So That All Students Benefit. A Nonprofit Approach for Equal Access!' (Khan Academy Blog, 14 March 2023) https://blog.khanacademy.org/harnessing-ai-so-that-all-students-benefit-a-nonprofit-approach-for-equal-access/ accessed 26 March 2023.

• Be My Eyes' Virtual Volunteer for vision assistance.[22]

Figure 2: The foundation model supply chain

UPSTREAM

WHAT: Gathers data at scale

WHO: Company collecting/ processing data

DATA LAYER

COMPUTE LAYER

WHAT: Provides processing power for training

WHO: Company providing computational resources

WHAT: Designs foundation model and trains foundation model using data and compute

WHO: Foundation model company

FOUNDATION MODEL DEVELOPER LAYER

HOST LAYER

WHAT: Hosts and makes models available to downstream users via download or API access

WHO: In-house at foundation model company or cloud service provider

WHAT: Builds and releases an application incorporating the foundation model or a fine-tuned version

WHO: Application developer (may be same as foundation model company or a third party)

APPLICATION LAYER

DOWNSTREAM

APPLICATION USER

WHAT: Uses application

WHO: End users (public. organisations, companies, etc.)

Note: This is one possible model (there will not always be a separate or single company at each layer)

---

22   Be My Eyes, 'Introducing Our Virtual Volunteer Tool for People Who Are Blind or Have Low Vision, Powered by OpenAI's GPT-4' (2023) https://www.bemyeyes.com/blog/introducing-be-my-eyes-virtual-volunteer accessed 26 March 2023.

Four ways models can be deployed

Currently, foundation models are made accessible to government clients and the public in four broad ways:

- **Systems built in-house**, e.g. development, training and deployment of a foundation model within government, either starting from scratch or fine-tuning an open-source model using government data.

- **Systems built for a customer**, e.g. a model procured from a third-party who fine-tunes it on behalf of government. The model may then be integrated into existing systems, e.g. a document search tool that sits on top of existing internal databases.

- **Systems relying on another application programming interface (API)**, such as integrating Open AI's GPT-4 API into a government service to enable services to provide large language model (LLM)-powered outputs. This means sharing input data with the foundation model provider, receiving output data back and feeding this back into the service, e.g. providing a response to a user query or generating an image for an internal drafting service.

- **Systems directly integrated into consumer tools**, e.g. image generation tools in Microsoft PowerPoint, automated editorial feedback in Google Docs, or ChatGPT. Here the foundation model is integrated into the backend of the consumer product and doesn't require technical understanding or work from the user. Integrating foundation models into existing software packages or making them available as subscription web services could mean they are accessed more informally by public sector officials.

Several key private sector firms also directly or indirectly collaborate with or pitch to government to roll out these tools.

OpenAI, Microsoft, Google and Anthropic offer developers and end users access to their foundation models via application programming interfaces (APIs). Through its API, OpenAI offers access to the GPT-3.5 and GPT-4 families of text-based foundation models, the DALL-E 2 image generation models, the Whisper-1 speech recognition model and more through its API.[23]

Other companies, like Hugging Face, instead offer access to open-source models, either via the Hugging Face API or downloaded and hosted directly on government servers.[24] All of these leading Western foundation model developers are US-based, or in the case of DeepMind owned by US-based company, Google.

---

23 OpenAI, 'OpenAI Platform' (2023) https://platform.openai.com accessed 1 August 2023.
24 Hugging Face, 'Expert Acceleration Program – Hugging Face' (2023) https://huggingface.co/support accessed 1 August 2023.

On this foundation, many companies are building customised products and services for clients. Among prominent actors in the UK context, Faculty, a UK-based company focusing on 'responsible AI' that has already partnered with OpenAI, and American data analytics company Palantir both have a long relationship with the government, supplying machine learning-based products and data infrastructure.

Numerous smaller firms also aim to provide foundation-model-based services to the UK government, such as domestic startup MantisNLP, which has built prototype chat interfaces for GOV.UK using foundation model APIs.

What is the broader public service and social context in which foundation models are being introduced?

The public sector faces many long- and short-term challenges, including, critically:

• recovery from the COVID-19 pandemic
• the cost-of-living crisis
• international pressures such as the changed relationship with the European Union following Brexit and the invasion of Ukraine impacting energy and food costs.[25]

Increased demand for services and falling real-terms budgets have led the public sector to seek efficiency and cost-cutting opportunities.[26] Many in local and central government see data-driven technologies as offering such opportunities.

However, harnessing data and technology has long been a challenge for the UK government. Outdated IT systems are common. Data architecture is fragmented, with information sitting in organisational silos. This makes it difficult to gather and summarise valuable information at speed or in a crisis; impedes holistic analysis, collaborative decision-making and coordinated service delivery; and creates costs and barriers to innovation. The civil service recognises these pressures but has not found solutions despite many digital transformation initiatives.[27]

---

25   Hannah White, 'Government in 2023: What Challenges Does Rishi Sunak Face?' (Institute for Government, 12 January 2023) https://www.instituteforgovernment.org.uk/publication/government-2023 accessed 4 July 2023.

26   Stuart Hoddinott, Matthew Fright and Thomas Pope, '"Austerity" in Public Services: Lessons from the 2010s' (Institute for Government 2022) https://www.instituteforgovernment.org.uk/sites/default/files/publications/austerity-public-services.pdf accessed 5 July 2023.

27   Daany Ajaib and others, 'Future Issues for Public Service Leaders' (2022) 4–5 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1093046/Future_Issues_for_Public_Service_Leaders.pdf accessed 4 July 2023.

In this context, thinktanks such as Demos, Reform and the Tony Blair Institute have called for public service provision to be rethought – from transactional to relational public services,[28] or reactive to proactive public services[29] – and for civil service reforms.[30]

Foundation models are not necessarily the only or right answer to public sector challenges. It is important to understand and define the problems first rather than rushing into technical solutions because they happen to have been developed at the moment a solution is needed.

> Government should critically compare foundation models to existing technical and policy approaches rather than see them as a default option. Are existing tools fit for a given purpose? Would they provide better value or better uphold public-sector principles?

Asking these questions prevents overzealous adoption of new technologies. Starting with people's needs fosters public-sector innovation that benefits society, rather than co-opting technology to drive change for its own sake.

---

28   'In this vision of public services, communities are given greater control over the way that services are delivered and users are treated as citizens, with something to offer and contribute to resolving the challenges they face [...] [relational public services consider] people holistically, looking at how they have got to where they are, the relationships and networks around them and treating them as a citizen whose view is to be respected, rather than tolerated.' See: Polly Curtis, Ben Glover and Andrew O'Brien, 'The Preventative State: Rebuilding Our Local, Social and Civic Foundations' (Demos 2023) https://demos.co.uk/research/the-preventative-state-rebuilding-our-local-social-and-civic-foundations/.

29   'Proactive public services (PPS) use existing data on individuals and businesses to help decide if they are eligible for a service and to trigger provision [...] [For example], in the Netherlands and the UK, annual tax returns are pre-filled with existing information (for example, from payroll records). Citizens only have to check and, if necessary, correct or add to the information before submitting the return.' See: Hilda Barasa and Alexander Iosad, 'What Are Proactive Public Services and Why Do We Need Them?' (11 October 2022) https://www.institute.global/insights/public-services/what-are-proactive-public-services-and-why-do-we-need-them accessed 5 July 2023.

30   Hannah White (n 25).

Current use and future application of foundation models in government and public services

This section highlights challenges identified during our research, such as skills gaps, leadership knowledge, navigating legislation and technical dependencies. Considering these limitations is essential for policymakers and public sector organisations seeking to use foundation models responsibly.

Informal usage

Leaders in local authorities noted that many staff members already informally use foundation-model-powered tools like ChatGPT and Bard in their work.[31] One leader shared an experience from a staff away day. They were surprised to discover many staff already using ChatGPT inside and outside of work.[32]

Similarly, by February 2023, the Department for Science, Innovation and Technology (DSIT) was reportedly being asked by other departments whether civil servants could use ChatGPT to help produce briefing materials. DSIT urged caution but did not rule it out.[33]

David Eaves, Associate Professor in Digital Government at the UCL Institute for Innovation and Public Purpose, likened this to a 'Dropbox moment'. Such moments have happened previously when staff began to use online tools like Dropbox and Wikipedia for work without formal introduction or monitoring by IT departments.[34] However, foundation models have been adopted much faster than previous tools. Leaders identified this as driven by a combination of:

- ChatGPT, Bard etc. being freely available and easily accessible
- the accessibility provided by natural language interfaces, which have made advanced models more user-friendly. For example, ChatGPT has made the GPT-3.5 model easily usable for a broad audience, not just data scientists and coders
- high awareness of these tools through social and traditional media.

31   London Office of Technology and Innovation Roundtable on Generative AI in Local Government (8 June 2023).

32   ibid.

33   Ben Gartside, 'Civil Servants Warned Not to Use AI Chatbots to Write Government Policies' (inews.co.uk, 21 February 2023) https://inews.co.uk/news/chatgpt-civil-servants-warned-ai-chatbots-government-2162929 accessed 30 May 2023.

34   Ada Lovelace Institute industry and civil society roundtable on the use of foundation models in the public sector (2023).

As discussed later, this rapid uptake has already prompted UK government departments to issue guidance on the use of publicly available foundation-model-powered tools and prohibit some use cases.[35]

Proposed use cases

Respondents both within public-sector product teams and from the private-sector companies Faculty, Palantir and Accenture told us that many public-sector organisations are enthusiastic about the promise of foundation models. However, adoption is still in initial stages. Beyond the informal use referred to above, participants were aware only of demos, prototypes and proofs of concept for foundation model use cases. They were not aware of any live system available to a large number of civil servants or the public.

According to participants, most proposed use cases focus on long-standing challenges facing public-sector organisations, like improving access to information across and between departments and public bodies and enhancing management of public enquiries. Foundation models are seen as a novel technological solution for tasks such as: [36]

- summarising complex reports
- translating documents
- analysing large datasets
- responding to public enquiries.

One area that elicited interest was using these systems for **internal knowledge management**. Participants felt foundation models could allow more efficient access to information spread across government – for example, by automatically aggregating policy documents, news articles and reports.

Some participants cautioned that poor-quality underlying data across government could lead to poor-quality outputs, limiting the usefulness of these systems. They emphasised the need to for the introduction of foundation models to happen alongside broader plans to improve data infrastructure.[37]

Another suggested application was **addressing public enquiries**. For example, departments could use foundation models to summarise public enquiries received through multiple channels. Data on

---

35   Central Digital and Data Office, 'Guidance to Civil Servants on Use of Generative AI' https://www.gov.uk/government/publications/guidance-to-civil-servants-on-use-of-generative-ai/guidance-to-civil-servants-on-use-of-generative-ai accessed 29 June 2023.

36   For a more detailed table of all the proposed use cases we heard from participants, and some more detailed hypothetical examples that were presented to participants ahead of the roundtable, please see Appendix 1.

37   The Office for Statistics Regulation covers this issue in its report 'Data Sharing and Linkage for the Public Good' (Office for Statistics Regulation 2023) https://osr.statisticsauthority.gov.uk/wp-content/uploads/2023/07/202307_office_statistics_regulation_data_sharing_linkage_report.pdf accessed 21 August 2023.

public enquiries is often fragmented, making it difficult for civil servants to see the whole picture. Foundation models could help categorise and distil data.

However, it was noted that success requires carefully considering what constitutes trustworthy information in this context. Just maximising accuracy is insufficient. For instance, existing data on public enquiries likely contains historical biases that could be inherited or amplified by foundation models. There are risks that certain groups may be unfairly profiled or misrepresented. Concrete steps must be taken to ensure fairness and avoid perpetuating biases. This might involve auditing data and models for discrimination and involving impacted communities.

Many participants suggested government should start with small-scale pilots focusing on discrete, low-risk use cases. For example, a foundation model could help **triage casewor**k by summarising PDF submissions and proposing categories for new cases. Specialists could then quickly evaluate if the system had assigned cases appropriately. This would provide a contained environment in which to evaluate performance.

Evaluations of pilots like this could inform use cases with greater potential benefit or harm. Participants indicated that such use cases would need continuous monitoring and oversight to ensure standards were upheld, even if they were determined to be low risk.

Representatives from industry described perceptions of these systems among public-sector clients as a 'magic' solution to any problem. They claimed that such clients had a 'fear of missing out' on an 'exciting' new technology, creating pressure for widespread adoption in the public sector.

However, participants across the board emphasised the need for realism. They acknowledged that the longstanding challenges of implementing data-driven systems in government will not disappear overnight. We explore these challenges in greater depth later in this section.

How might foundation model use cases be implemented and rolled out within government?

Civil society and industry participants were split on what balance government should strike between procurement and internal development. Government could buy in external foundation-model-powered tools, it could develop its own foundation models or foundation-model-powered applications, or it could do both.[38]

---

38   Ada Lovelace Institute industry and civil society roundtable on the use of foundation models in the public sector (2023).

Private firm Faculty has partnered with OpenAI to deploy foundation models to clients, including governments. For example, it has worked with NHS England to create synthetic doctors' notes to evaluate privacy risks in NHS free text data, e.g. doctors' written comments.[39]

This approach is promoted as giving government faster access to some cutting-edge models. However, there is a risk of lack of internal government capacity to understand and develop tools. It could also leave government reliant on private-sector providers in future.

Participants also noted a potential lack of alignment between applications developed for a wider range of clients and the needs of the the public sector. In particular, public-sector clients:

• are more likely to deal with highly sensitive data
• have higher standards of robustness
• require higher levels of transparency and explainability in important decisions around welfare, healthcare, education and other public services.

David Eaves has suggested that the public sector should be 'fast followers', adopting established technologies and practices rather than always trying to use the very latest options. He points to the first few years of the Government Digital Service's modernisation programme an example. Being a fast follower does not necessarily require procuring external tools and could help ensure that modernisation happens steadily, using proven methods.

Other participants raised the prospect of government taking a more proactive role in developing foundation models.[40] This would involve developing internal capacity to train or fine-tune foundation models rather than relying on access to APIs or third-party vendors.

This idea has been discussed by academics and think tanks across the political spectrum, often referred to as a **'sovereign LLM [large language model]' or 'BritGPT'**.[41] Under this model, the UK government would own rather than access LLMs (and in future other forms of foundation models). Advocates of this approach argue that it would enhance public service delivery using a tailored and safe LLM. Many proposed versions involve collaboration rather than competition with the private sector.

---

39   'OpenAI Asks Faculty to Help Businesses Adopt Generative AI' (Faculty, 13 March 2023) https://faculty.ai/blog/openai-asks-faculty-to-help-businesses-adopt-generative-ai/ accessed 5 July 2023.

40   Ada Lovelace Institute industry and civil society roundtable on the use of foundation models in the public sector (2023).

41   Haydn Belfield, 'Great British Cloud And BritGPT: The UK's AI Industrial Strategy Must Play To Our Strengths' (Labour for the Long Term 2023) 6 https://www.labourlongterm.org/briefings/great-british-cloud-and-britgpt-the-uks-ai-industrial-strategy-must-play-to-our-strengths. Shabbir Merali and Ali Merali, 'The Generative AI Revolution' (Onward 2023) 40 https://www.ukonward.com/reports/the-generative-ai-revolution/.

Industry and civil society participants highlighted that this **societal-led technology does not come for free**.[42] Many proposals for public-sector foundation model development anticipate costs of between £200 million and £1 billion over the next couple of years. The UK government was seen to be at the beginning of this journey. It was deemed extremely important to properly engage the public on any 'BritGPT', 'Sovereign LLM' or other significant investment in public-sector AI development.

Either way, as an attendee at a roundtable hosted by the Centre for Data Ethics and Innovation (CDEI) noted: 'government is better off specialising and innovating in user-centric design and deployment of models in ways that are sensitive to ethical concerns - as proof that this can be done and as a benchmark that industry can be measured against.'[43]

> The Ada Lovelace Institute believes it is unlikely that training models to replicate or compete with foundation models such as GPT-4 would unlock significant benefits for people and society at proportionate cost.[44]

However, we believe it would be valuable for government to explore how public support could facilitate development of AI technologies and applications that are not currently well-served by market trends, including those related to public sector foundation model deployment.

Challenges in deploying foundation models in the public sector

Participants in the roundtables identified several challenges in deploying foundation models in the public sector.

Cybersecurity

Public sector organisations often manage personal, special category and other sensitive data and are common targets for cyber attacks.

---

42   Ada Lovelace Institute industry and civil society roundtable on the use of foundation models in the public sector (2023).
43   'Centre for Data Ethics and Innovation roundtable on use of foundation models in the public sector' (8 June 2023).
44   Ada Lovelace Institute, Regulating AI in the UK (2023) https://www.adalovelaceinstitute.org/report/regulating-ai-in-the-uk/ accessed 1 August 2023.

Malicious actors could exploit AI systems, including LLMs and other foundation models, if systems are not properly secured. This could have serious implications for national security and public trust. Plugins and other tools that allow foundation models to interact with the wider web are particular points of insecurity.[45]

These systems could also violate internal permission structures, revealing classified information. Advancements in AI security could reduce these risks over time, but more-capable foundation models are not necessarily more secure.

Off-the-shelf solutions could provide access to cutting-edge models. However, this would leave the cybersecurity of public sector foundation model applications dependent on the security of private providers. If the public sector is willing to invest the necessary resources, a self-hosted foundation model could offer greater control over the robustness and security of applications.

Skills and careers

Roundtable participants raised the challenge of recruiting and retaining skilled data professionals in the public sector, especially outside London.[46] This is primarily due to the sector's competitive nature and the concentration of talent in specific geographical locations.

Furthermore, there were concerns about the potential displacement of entry-level jobs and the consequent impact on career pathways. However, there is a perceived opportunity to use AI to enhance areas which are already losing staff, such as social care, potentially making future recruitment easier if the role becomes less monotonous, more productive or better paid.

Leadership understanding

Roundtable participants emphasised the importance of decision-makers having a robust understanding of the technologies they are deploying. This is especially true for public-sector leaders, who are often non-technical or removed from the day-to-day process of working with the technology.

The overlapping terms used to refer to these systems (foundational model, generative AI, general-purpose AI, frontier AI, etc.) can increase confusion. Clear agreement is needed on terms and their meanings, at least within government.

---

45   Ada Lovelace Institute government roundtable on the use of foundation models in the public sector (2023).
46   ibid.

Some participants noted that a few London councils have a Chief Data Officer (CDO) in their management team. These CDOs provide explicit leadership at a senior level on data-related topics, and they could play a significant role in successfully implementing foundation models.

Navigating legislation, regulation and guidelines

Participants told us that a lack of clear guidance could hinder the deployment of foundation models.

Developers, deployers, and users must follow existing regulations. However, current UK regulation in this area is fragmented. This includes 'horizontal' cross-cutting frameworks, such as human rights, equalities and data protection law, and 'vertical' domain-specific regulation, like the regime for medical devices.

Guidance is slowly emerging on interpreting relevant laws and regulations in the context of AI, including foundation models, but the regulatory environment remains complex and lacks coherence.[47] We discuss this below in the section on 'Principles and existing governance mechanisms'.

Engaging legal teams from the beginning was seen as essential, to help mitigate potential data protection issues. Take the example of using foundation models to analyse social media posts during immigration casework. In this high-risk use case, tighter controls may be needed to ensure data privacy and security standards. Legal teams could provide assurance to product teams and make sure they take proactive measures to prevent legal breaches.

Participants from product teams told us that building AI literacy across their organisations was an ongoing challenge. Professionals in data, legal and data protection roles should receive training and support to improve their understanding of AI, help them keep pace with new developments, and support them to provide adequate oversight and guidance on the use of foundation models.

---

47   See for examples: ICO (Information Commissioner's Office), 'Artificial Intelligence' (19 May 2023)
      https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/ accessed 1 August 2023;
      Google, 'Google AI Principles' (Google AI) https://ai.google/responsibility/principles/ accessed 1 August 2023; TUC,
      'Work and the AI Revolution' (25 March 2021) https://www.tuc.org.uk/AImanifesto accessed 1 August 2023. Equity, 'Equity AI Toolkit'
      (Equity) https://www.equity.org.uk/advice-and-support/know-your-rights/ai-toolkit accessed 1 August 2023"plainCitation":
      "Equity, 'Equity AI Toolkit' (Equity; Cabinet Office, 'Guidance to Civil Servants on Use of Generative AI' (GOV.UK, 2023)
      https://www.gov.uk/government/publications/guidance-to-civil-servants-on-use-of-generative-ai/guidance-to-civil-servants-on-use-
      of-generative-ai accessed 1 August 2023.

Novelty bias

There is a risk that foundation models are used because they are a new 'shiny' technology everyone is talking about, not because they are the best tool or system for the challenge.

This pattern has been seen before with other innovations. Government participants referred to the hype around blockchain as an example. Being influenced to adopt a technology because it is new and exciting can lead to poor outcomes and waste financial and human resources.

Frequent changes in senior leadership priorities were also seen as a barrier. These changes make it difficult for product teams to focus on solving substantive problems. A tendency to take up digital government trends often leads to short-lived pilots. These pilots are not always taken forward when they succeed, or those that fail may still gain momentum and become a costly norm. Consistently evaluating new tools against concrete problems and then committing to longer-term iterated projects will lead to more prudent investments.

The role of the Frontier AI Taskforce

Many participants cited the Frontier AI Taskforce (formerly the Foundation Model Taskforce) as an important initiative. They believed it would be important in coordinating the use of foundation models across the public sector and resolving some of the challenges described above.

Participants wanted to see the Taskforce facilitate pilot deployments in contained environments. This would allow government to gain hands-on experience while evaluating performance and risks. This could then inform guidelines and best practices that could be rolled out for foundation models across government. This development of guidelines and best practices would include contributing to ongoing work on AI governance, standards and practices by existing government teams such as the Office for AI (OAI), the CDEI, and the Central Digital and Data Office (CDDO).

Participants hoped the Taskforce would also help to explore responsible applications aligned with public sector values. By bringing together activities under a central body, the Taskforce could drive progress that individual departments might struggle to achieve. It could also play an important role in addressing cross-cutting challenges like skills development, data access, technical infrastructure and legal uncertainties.

However, participants emphasised the need for transparency about the Taskforce's activities. Simply asserting that these technologies will modernise government is insufficient. People expect public-sector AI use to be evidence-based and to uphold stringent standards. The Taskforce should proactively communicate how it would assess the fairness, accountability and public benefit of proposed use cases. It should also make meaningful public engagement part of its

governance structures and objectives. The priorities and concerns of the public and not just those of government institutions should shape its work.

Civil society organisations– including consumer groups, trade unions, advice-giving bodies such as Citizens Advice, and charities representing vulnerable people and those with protected characteristics – also have a crucial role. Yet initial Government communications on the Taskforce failed to mention civil society expertise or participation by individuals and groups affected by AI. We would welcome a Government commitment to meaningful involvement of these groups.[48]

---

48   Ada Lovelace Institute (n 44).

Risks of foundation models in the public sector

This section examines the various risks and challenges to developing, procuring and deploying large language models (LLMs) and foundation models in the public sector. It applies to foundation models the emerging taxonomy of the risks of language models. We look at bias, privacy violations, security threats, environmental impacts and workforce harms and analyse how they may manifest in public sector use cases.

Taxonomies of risk and harm

There are numerous risks and potential harms common to any use of algorithmic systems.[49] There are also concerns about the development of more capable AI in general, including malicious uses in bioterrorism and surveillance, arms races, accidents that result in systemic shocks or critical infrastructure failure, and deceptive AI systems.[50] All of these risks may pose challenges in the present or near-term deployment of foundation models in a public-sector context.

Researchers at DeepMind have developed a taxonomy of ethical and social risks from language models, which can be generalised as a taxonomy of risks from present and near-term foundation models in general.[51] The researchers identified **six broad categories of risk:**

- **Discrimination, hate speech and exclusion:** arising from model outputs producing discriminatory and exclusionary content.
- **Information hazards:** arising from model outputs leaking or inferring sensitive information.
- **Misinformation harms:** arising from model outputs producing false or misleading information.
- **Malicious uses:** arising from actors using foundation models to intentionally cause harm.
- **Human-computer interaction harms:** arising from users overly trusting the foundation model, or treating them as human like.

---

49   Renee Shelby and others, 'Identifying Sociotechnical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction' (arXiv, 8 February 2023) http://arxiv.org/abs/2210.05791 accessed 27 March 2023."plainCitation":"Renee Shelby and others, 'Identifying Sociotechnical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction' (arXiv, 8 February 2023

50   Dan Hendrycks, Mantas Mazeika and Thomas Woodside, 'An Overview of Catastrophic AI Risks' (arXiv, 26 June 2023) http://arxiv.org/abs/2306.12001 accessed 4 July 2023.

51   Laura Weidinger and others, 'Taxonomy of Risks Posed by Language Models', 2022 ACM Conference on Fairness, Accountability, and Transparency (Association for Computing Machinery 2022) https://doi.org/10.1145/3531146.3533088 accessed 30 January 2023."plainCitation":"Laura Weidinger and others, 'Taxonomy of Risks Posed by Language Models', 2022 ACM Conference on Fairness, Accountability, and Transparency (Association for Computing Machinery 2022

- **Automation, access and environmental harms:** arising from the environmental or downstream economic impacts of the foundation model.

Many of these concerns are best dealt with by the upstream providers of foundation models at the training stage (e.g. through dataset cleaning, instruction fine-tuning or reinforcement learning from feedback).[52] But public-sector actors need to be aware of them if developing their own foundation models, or to ask questions about them when procuring and implementing external foundation models.

For example, when procuring or developing a summarisation tool, public sector users should ask how issues like gender or racial bias in text outputs are being addressed through training data selection and model fine-tuning. Or when deploying a chatbot for public enquiries, they should ensure that process of using data to prompt the underlying large language model does not violate privacy rights, such as by sharing data with a private provider with poor cybersecurity.

By factoring these risks into procurement requirements, pilot testing and ongoing monitoring, public sector users can pressure external providers to take steps to minimise downstream harms from deployment within government contexts.

---

52   ibid 222.

Ethical and social risks from foundation models

Risk category	Harm	Description
Discrimination, hate speech and exclusion	Social stereotypes and unfair discrimination	Foundation models are often trained on publicly available text, images and videos from across the internet, including Reddit, Wikipedia, and other websites. This data often reflects historical patterns of systemic injustice and inequality. As a result, the model can learn and reproduce demeaning language and harmful stereotypes about marginalised groups in society.[53] In a public-sector context, this could result in reproducing stereotypes in draft documents. For example, racially minoritised groups might be disproportionately represented in case studies about violent crime, or inappropriately gendered language might be used in relation to healthcare.
	Hate speech and offensive content	Foundation models 'may generate [content] that includes profanities, identity attacks, insults, threats, that incites violence, or that causes justified offence as such [content] is prominent online. This [content] risks causing offence, psychological harm, and inciting hate or violence.'[54] Many models have been fine-tuned and content-moderated to avoid this, but it can still occur during use. This could be particularly harmful in an educational context, leading to an unsafe learning environment and psychological harm to students. Offensive content entering into internal government communications such as draft emails or other communications could cause also problems, such as harming workplace relations or damage the professional integrity of the public sector.
	Worse performance for some languages and social groups	Foundation models generally perform less well on text and voice tasks for less widely spoken languages, dialects and accents. [55] This is because they are typically trained on only a few languages (often overwhelmingly English). This is partly because of a lack of training data; for less widely spoken languages, there may be little digital content available and datasets may not have been assembled. The excluded languages, dialects and accents are more likely to be used by marginalised or excluded communities, especially refugees and asylum seekers, who are already at a heightened risk of harm. As foundation models are integrated into public services, this may lead to unequal access, poorer-quality services and more negative experiences of those services by people those first language is not English.

---

53   ibid 216.

54   ibid.

55   ibid 217.

Risk category	Harm	Description
Information hazards	Compromising privacy by leaking sensitive information	Foundation models can 'remember' private information from data they are trained on, leading to an inherent risk of privacy violation and confidentiality breaches. [56] This risk is particularly acute when models are fine-tuned on internal government data and data could be leaked across governmental departments.

AI systems, especially foundation models, often require access to extensive amounts of data for training and deployment. Managing this requires compliance with all relevant data protection and privacy laws. It is also necessary to navigate the complexities of access control during all stages of development and deployment. This poses a significant operational challenge to maintaining data privacy and confidentiality.

Government participants highlighted a need for UK-based data centres for applications that process sensitive and classified information, [57] to reduce the risk of cross-border data leaks and ensure greater control and protection of confidential data. |
| | Compromising privacy or security by correctly inferring sensitive information | Foundation models can compromise privacy even without personal data being used to train the model, by making accurate guesses about sensitive personal details such as a person's sexuality, religion or gender identity. [58] This can happen based on how a person writes or speaks to the model.

For example, a public-sector organisation might use foundation models to assist in screening job applications or in performance evaluations. A sufficiently capable foundation model could learn patterns associated with protected characteristics and could infer protected traits of applicants/employees based on the content of their application or professional communications. It might infer an applicant's religious beliefs, political affiliations or sexual orientation from their activities, clubs they belong to or even the phrasing of their cover letter. The applicant would not have intentionally disclosed this information, so its inference is a privacy violation. Such inferences count as 'special category' data under the UK General Data Protection Regulation (GDPR) and similar data protection regimes.

Equally, there are risks if such inferences are invalid but believed to be correct. For example, a false inference about a person's sexual orientation or gender identity may lead to discrimination or emotional harm when shared or acted on. Or a model could inaccurately infer an employee's political or religious beliefs based on their communications and this could subtly bias their manager's performance review. |

56   ibid.
57   Ada Lovelace Institute government roundtable on the use of foundation models in the public sector (n 45).
58   Laura Weidinger and others (n 51) 218.

Risk category	Harm	Description
Misinformation harms	Disseminating false or misleading information, potentially leading to material harms	A core concern about the current limitations of systems is 'hallucinations' – where an AI system provides a confident answer, but one that bears no relation to reality. Foundation models lack common sense: they simply predict the most likely word, pixel or other item based on training data. They may be fine-tuned to appear 'helpful, harmless and honest', but they have no true understanding. They have no way to know if their outputs are accurate or fabricated. As a result, they can confidently produce misleading, biased or false information. Because the systems sound convincing, people may believe and act on these falsehoods. The temporal limitations of these technologies present a related challenge. Systems have a cut-off point for training data: for example, OpenAI's GPT-3 training data extended until September 2021. Most existing foundation models are not continually retrained with updated information and cannot draw on up-to-date external data sources. This increases the chance that they will generate inaccurate outputs such as incorrect information about current leaders or government policies. Misleading information could lead to people to waste time applying for benefits they are not entitled to,[59] or to fail to claim benefits they are entitled to. And a poorly developed chatbot could actively encourage unethical behaviour. For example, a tax return completion assistant intended to help people minimise their tax burden could advise people to claim dubious deductions or even encourage tax evasion.

---

59   For a more detailed exploration of this scenario, and the relevant legal and regulatory protections under current UK law, see:
     Alex Lawrence-Archer and Ravi Naik, 'Effective Protection against AI Harms' (AWO 2023) 54–64
     https://www.awo.agency/files/AWO%20Analysis%20-%20Effective%20Protection%20against%20AI%20Harms.pdf
     accessed 16 August 2023.

Risk category	Harm	Description
Malicious uses While it is likely that most users would behave appropriately, foundation models can be used inappropriately or maliciously. Preparing for how to manage such an eventuality is important.	Making disinformation cheaper and more effective	Foundation-model-assisted content generation potentially offers a lower-cost way of creating disinformation at scale. [60] Disinformation campaigns may be used to mislead the public or shape public opinion on a particular topic, such as opposing or supporting a new government policy. Large language models may be used to create large volumes of synthetic responses to public consultations, skewing perceptions of majority opinion and creating additional work for civil servants.
	Assisting code generation for cyber security threats	Assisting coding tools like GitHub Co-Pilot, which are based on OpenAI's GPT models, may make it easier and cheaper to develop malware, and to make it more adaptable in order to evade detection.[61] Public services are already being hit by malware attacks. The NHS WannaCry ransomware attack in 2017 is one example,[62] and Gloucester City Council services were crippled for hours after an attack by Russia-linked hackers.[63] These attacks disrupt services, compromise sensitive and personal data, and burden public bodies with high recovery costs . As AI-driven development becomes more mainstream, a rise in more sophisticated, AI-generated cyber threats is plausible. Public services need advanced cybersecurity measures, including AI-powered threat detection and response systems that are a match for sophisticated AI-enhanced malware attacks. The cost of this needs to be taken into account.
	Facilitating fraud, scams and targeted manipulation	Foundation models could be used by criminals to make frauds and scams more targeted.[64] For example, they could be fine-tuned based on speech data to more accurately impersonate an individual for the purpose of identity theft. They could also be used to cause harm at greater scale, for example by generating more personalised, compelling email scam text or by maintaining an extended conversation with a victim. Generated content could also be fraudulently presented as a person's own work, for example to cheat on an exam. There are specific risks to the public sector. Phishing scams may target civil servants, tricking them into revealing sensitive information or granting access to secure systems. Criminals could also use these technologies to craft more convincing false claims or falsify documents, enabling welfare, tax and government fraud. Impersonation of staff could cause many problems, from misinformation to damaging the reputation of public institutions.

60   Laura Weidinger and others (n 51) 219.

61   ibid.

62   Investigation: WannaCry cyber attack and the NHS - NAO report, (no date), https://www.nao.org.uk/reports/investigation-wannacry-cyber-attack-and-the-nhs/ (Accessed: 7 July 2023) *NHS ransomware attack: what happened and how bad is it? | Cybercrime | The Guardian*, (no date), https://www.theguardian.com/technology/2022/aug/11/nhs-ransomware-attack-what-happened-and-how-bad-is-it (Accessed: 7 July 2023);

63   'Gloucester Cyber Attack: A Year since Hackers Disrupted Vital Services for Thousands of Citizens - Gloucestershire Live' https://www.gloucestershirelive.co.uk/news/gloucester-news/gloucester-cyber-attack-year-hackers-7947958 accessed 7 July 2023.

64   Laura Weidinger and others (n 51) 219.

Risk category	Harm	Description
Human–computer interaction harms	Promoting harmful stereotypes by implying gender or ethnic identity	Foundation model applications could perpetuate stereotypes by using identify markers in outputs. For example, if an application refers to itself as 'she' or by a gendered name (such as Alexa) this may fuel the expectation that women naturally take on assistant roles. In a public-sector context, if a foundation model used for drafting policy documents defaults to male pronouns and gendered language, this could contribute to underrepresentation of women and non-binary people in public life.
	Anthropomorphising systems and automation bias can lead to overreliance or unsafe use	Overreliance on these systems is a significant concern, particularly where decisions can have far-reaching implications – not least for the wellbeing of individuals. Foundation models can generate natural language and consistent images and videos for chatbots or other interactive applications. If people see these 'agents' as human-like they may place undue confidence in the applications.[65] More capable foundation models could also lead to automation bias – where users trust and defer to automated systems even when their decisions or outputs are flawed. Even 'simple' tasks like summarisation may lead to filter bubbles or biased documents based on the quality of the prompt or the system and not on the knowledge and expertise of stakeholders. Furthermore, when an automated system's choices or actions are deferred to, this can further blur accountability within the system and how affected individuals can seek redress.
	Avenues for exploiting user trust and accessing more private information	Foundation models can be used to create natural-seeming conversational agents. The resulting natural conversation flow and perceived anonymity can make users feel more comfortable.[66] They could therefore be willing to disclose sensitive and private information, including about personal hardships, mental health issues, family problems or opinions on government policies. If a public body recorded and analysed this data without appropriate privacy measures, this could lead to privacy violations and misuse of sensitive information, such as making intrusive inferences about the individual's lifestyle or condition. This could lead to biased decision-making, for example in welfare provision. If the public body then shared that information, e.g. with law enforcement or immigration services, and these other services used it against the individual, this would harm the individual and make others less willing to engage with government services in future. This is particularly true of people from vulnerable or marginalised groups.

65   ibid 220.
66   ibid.

Risk category	Harm	Description
Automation, access and environmental harms	Environmental harms	Developing and deploying foundation models is likely to take significant amounts of energy. Given the UK's current methods of energy production, this could have a detrimental impact on the environment. Foundation models can have an environmental impact through:[67]
		• the direct energy required to train and use the model
		• indirect energy use for foundation-model-powered applications
		• the resource extraction required to create the hardware and computing chips to train and run the models
		• systemic changes in human behaviour due to foundation model applications.
		If Government is to meet its net zero targets and other environmental commitments, it needs to understand and tackle the impact of foundation models.[68] The energy and resource extraction required may be an acceptable cost to access the benefits of previously discussed use cases. And if foundation model applications improve efficiency, this could mean less net energy consumption and resource extraction. Still, Government should assess the trade-offs for a given application, and whether it meets their existing sustainability commitments.
		The Climate Change Committee (CCC) has recommended that the UK Government extend climate skills training across the public sector.[69] This should include an understanding of the environmental externalities and trade-offs of AI deployment.

67   ibid 220–221.

68   'Greening Government Commitments 2021 to 2025' (Department for Environment Food & Rural Affairs 2022) https://www.gov.uk/government/publications/greening-government-commitments-2021-to-2025/greening-government-commitments-2021-to-2025 accessed 7 July 2023. 'Greening Government: ICT and Digital Services Strategy 2020-2025' (Department for Environment Food & Rural Affairs 2020) https://www.gov.uk/government/publications/greening-government-ict-and-digital-services-strategy-2020-2025/greening-government-ict-and-digital-services-strategy-2020-2025 accessed 7 July 2023.

69   '2023 Progress Report to Parliament' (Climate Change Committee 2023) locs R2022-144 https://www.theccc.org.uk/publication/2023-progress-report-to-parliament/ accessed 21 August 2023.

Risk category	Harm	Description
	Negative effects on labour rights and job quality	The commitments of successive UK Governments on labour rights are relevant both to how foundation-model-based systems are trained and to how and where they are used.[70]

Public-sector workers, particularly in frontline and administrative roles, are already experiencing increasing workloads, staffing shortages and declining budgets, affecting morale and job quality. Staff report an intensified pace of work, monotonous tasks and loss of autonomy. Similarly, automation in warehouses and the gig economy has tightened control over workers and reduced human collaboration.

Foundation models risk exacerbating these concerns. For example, by invasively monitoring staff behaviour through sentiment and topic analysis of private conversations. Some public-sector jobs could be displaced as systems become capable of automating tasks like responding to public enquiries or generating illustrations. People may be relegated to monitoring model outputs rather than applying creativity.[71]

Labour violations are also possible further up the supply chain – for example, data labellers may be exploited in the training of systems. It may be difficult for government to assess conditions due to poor-quality, incomplete or non-existent data on the supply chain for foundation models.

The Ada Lovelace Institute has called for standardised reporting on foundation model supply chains to be part of mandatory transparency requirements for developers.[72] |

70   Department for Business, Energy & Industrial Strategy, (2018), Good Work Plan, Policy Paper, https://www.gov.uk/government/publications/good-work-plan/good-work-plan (Accessed: 1 August 2023) Powell and Ferguson, (2020), 'Employment law in the modern economy: The Good Work Plan', https://commonslibrary.parliament.uk/employment-law-in-the-modern-economy-the-good-work-plan/ (Accessed: 1 August 2023);
71   Laura Weidinger and others (n 51) 221.
72   Ada Lovelace Institute (n 44).

Risk category	Harm	Description
	Disparate access to benefits due to hardware, software, skill constraints	Introducing foundation models in public-facing services could lead to inequality or discrimination for people and groups who cannot access the internet or digital services.[73] In previous work on contact tracing apps and vaccine passports during the COVID-19 pandemic, the Ada Lovelace Institute found that:[74] • nearly a fifth (19%) of respondents did not have access to a smartphone • 14% did not have access to the internet • 8% had neither a smartphone or access to the internet. This was particularly acute for disabled people, those on low incomes (less than £20,000) and those older than 65. We also found significant disparities in awareness, knowledge and comfort using digital health apps for disabled people and those on low incomes, compared to the general population. These technologies are not directly equivalent to foundation models, but these findings still underline the need to consider disparities during the rollout of any new technology, particularly in the public sector, which should serve everyone in society. Where necessary, non-digital services accessible to the digitally excluded should be funded alongside investment in foundation models.

---

73   Laura Weidinger and others (n 51) 221–222.

74   Ada Lovelace Institute, The data divide (2021) https://www.adalovelaceinstitute.org/survey/data-divide/ accessed 6 April 2021.

Principles and existing governance mechanisms

The risks and challenges discussed above make it clear that responsible governance and oversight are essential to any use of foundation models.

This section outlines some principles proposed for assessing whether public-sector use cases are appropriate and aligned with public-sector values. It then summarises existing governance mechanisms highlighted by roundtable participants, like data ethics frameworks and impact assessments, that provide a starting point for the governance of foundation models.

We then discuss a survey of existing legislation, regulation and guidance conducted by the AI Law Consultancy. Finally, we look at existing public sector guidance on foundation model use.

Principles for assessing use cases

The UK has seven principles for standards in public life (as known as the Nolan principles) which apply to anyone who holds public office, including politicians, civil servants and other public-sector staff:[75]

1. Selflessness
2. Integrity
3. Objectivity
4. Accountability
5. Openness
6. Honesty
7. Leadership

These principles provide a long-established and widely accepted standard that government has decided it should be held to. They are therefore a solid foundation for the governance of any rollout of foundation models in the public sector. Rather than reinventing the wheel, attention should be

---

75   'The Seven Principles of Public Life'
     https://www.gov.uk/government/publications/the-7-principles-of-public-life/the-7-principles-of-public-life--2
     accessed 25 August 2023.

turned to how foundation models can comply with these established standards.[76]

In 2020, the Committee on Standards in Public Life argued that all seven principles are relevant to AI use in public-service delivery but that AI posed a particular challenge to the principles of openness, accountability and objectivity.[77]

The Committee has requested information from public bodies and regulators about use of AI in decision-making and what governance arrangements are in place.[78] Its findings will likely provide a systematic account of how ready public bodies and regulators are to apply the Nolan principles to foundation models. We expect readiness to vary considerably across the sector, based on our roundtables and commissioned AI Law Consultancy research, discussed later in this section

Roundtable participants emphasised that the public sector should not be viewed as a monolithic entity. The complexities and nuances of central and local government, the NHS, the education system, and other parts of the public sector need to be acknowledged. What is appropriate in one setting may be problematic when applied to another area of government.

For example, a model used to help civil servants search internal documents in confidential government databases poses fewer direct risks to the public than one that interacts with individuals to provide information on government services. Internal search applications risk leaking sensitive data or perpetuating biases, but these can be corrected through internal training, guidance and human oversight. The potential harms are lower than those of a public-facing chatbot providing misleading advice that results in people not receiving benefits or misinterpreting legal obligations.[79]

Oversight and understanding of these technologies vary substantially between different bodies. For example, in healthcare, the Medicines and Healthcare products Regulatory Agency (MHRA) has said that 'LLMs [...] developed for, or adapted, modified or directed toward specifically medical

76   Researchers at the Council of the European Union General Secretariat have similarly suggested using the broadly equivalent eight European principles of public administration as a standard by which to judge the possible introduction of LLMs (and other foundation models) into the public sector. Analysis and Research Team, 'ChatGPT in the Public Sector – Overhyped or Overlooked?' (Council of the European Union General Secretariat 2023) 10 https://www.consilium.europa.eu/media/63818/art-paper-chatgpt-in-the-public-sector-overhyped-or-overlooked-24-april-2023_ext.pdf accessed 24 May 2023.

77   Jonathan Evans, 'AI Brings Potentially Huge Benefits to the Public Sector but We Need Clear Standards and Greater Transparency - Committee on Standards in Public Life' (25 May 2023) https://cspl.blog.gov.uk/2023/05/25/ai-brings-potentially-huge-benefits-to-the-public-sector-but-we-need-clear-standards-and-greater-transparency/ accessed 29 June 2023.

78   Jonathan Evans, 'Letter to Public Bodies on Artificial Intelligence and Public Standards Report Follow Up' (4 July 2023) https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1168366/Lord_Evans_letter_to_departments_and_public_bodies_on_AI.pdf accessed 10 July 2023. Jonathan Evans, 'Letter to Regulators on Artificial Intelligence and Public Standards Report Follow Up' (4 July 2023) https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1168358/Lord_Evans_letter_to_regulators_on_AI.pdf accessed 10 July 2023.

79   For more on this topic, see the analysis the Ada Lovelace Institute commissioned from AWO, analysing the UK's current legal regime to test how effective it is at protecting individuals from AI harms, including a hypothetical concerned a Chatbot used by a government agency to provide benefit advice which is sometimes incorrect. Lawrence-Archer and Naik (n 59) 54.

purposes', or that can be used for a medical purpose according to the developer, are likely to qualify as medical devices.[80] Developers of such LLMs must have evidence that they are safe under normal conditions of use and perform as intended, as well as having to comply with other requirements of medical device regulation.

Other regulatory authorities have yet to make even a provisional statement on the use of foundation models, LLMs and similar.[81] We further explore these differences later in this section.

Efficiency

Roundtable participants raised concerns that the adoption of foundation models was often framed around efficiency gains.[82] Participants noted a tendency to focus on optimising existing workflows. This was seen as more likely to make existing policies and practices more efficient, not to enable a broader reimagining of how services might work.

For example, using foundation models to optimise existing welfare practices could lead to deployers intensifying existing punitive practices rather than considering other strategies, such as using AI to find and contact people who are entitled to but do not receive particular benefits.[83] This narrow focus could lead to a risk of bias towards targeted and means-tested approaches in public services. This may not always be the appropriate response to a policy problem.

Participants suggested that foundation model applications should instead enable new approaches. This should align with a vision of societal outcomes and promoting social good and not simply replace jobs and automate existing workflows. Some suggested that if AI displaces call centre workers, government could redirect them to roles that enhance community engagement (rather than making them redundant, as is likely under the current UK public-sector model of trying to do more with less).

Counterfactuals

In implementing foundation models, civil servants should carefully consider the counterfactuals. This involves comparing the proposed model use case with other, more mature alternatives to question whether existing tools might be more effective or provide better value for money. This

---

80   Johan Ordish, 'Large Language Models and Software as a Medical Device' (MedRegs, 3 March 2023)
     https://medregs.blog.gov.uk/2023/03/03/large-language-models-and-software-as-a-medical-device/ accessed 6 July 2023.
81   AI Law Consultancy commissioned research.
82   Ada Lovelace Institute industry and civil society roundtable on the use of foundation models in the public sector.
83   This application itself could raise concerns. Some of those not accessing welfare may be choosing not to and may not wish to be found or engage with a state bureaucracy they feel is hostile to them.

evaluation should be guided by the principles of public life, with a critical eye towards efficiency.

Determining whether a foundation model use case is valuable may need demonstrations, proofs of concept or pilot tests. By making these counterfactual comparisons, civil servants can make informed, sensible choices about implementing foundation models. This helps to strike a balance between embracing new technology and making the most of existing tools.

Existing legislation, regulation and guidance

What we heard from our roundtables about existing regulation and guidance

Roundtable participants representing government departments told us that many parts of government already have strong governance around data ethics. This, alongside the Nolan Principles, could serve as a foundation for governance of foundation model use.[84] This optimism was tempered by concerns that existing structures might not provide sufficiently specific guidance for deploying foundation models.

Data Protection Impact Assessments (DPIAs) and Equality Impact Assessments were viewed as crucial baseline tools for risk management in deploying any algorithmic system.[85] This was because the risk associated with AI was seen as highly context-dependent, For example, autogenerating emails for internal communication is very different to doing so for patients or residents. Different scenarios need case-by-case risk assessment.

In this context, it is concerning that the Data Protection and Digital Information Bill (DPDI Bill), currently before Parliament, removes the obligation on data controllers to carry out DPIAs when high-risk processing is being carried out, as well as other safeguards vital to protecting people and communities from AI-related harms.

Roundtable participants were concerned that these amendments to regulation could lead to governance based on opinion rather than data-driven insights. The Ada Lovelace Institute believes that collectively, the changes in the DPDI Bill risk undermining the safe deployment of AI and that the proposed changes should be reconsidered.[86]

---

84   Ada Lovelace Institute government roundtable on the use of foundation models in the public sector.

85   ibid.

86   Ada Lovelace Institute (n 44).

A survey of existing legalisation, regulation and guidance

We wanted a more systematic view of the legislative, regulatory and guidance landscape for foundation models, and their application in the UK public sector. We commissioned AI Law Consultancy to survey, between 27 June and 19 July 2023, how foundation models, and related emerging AI concepts, are currently addressed in UK legislation, regulation and public guidance.[87] These concepts include:

- Foundation models
- General-purpose AI GPAI)
- Generative AI
- Large language models (LLMs)
- Artificial general intelligence (AGI)
- Frontier models.

Legislation

While various primary and secondary sources of legislation define AI in fields such as export control, the survey found that:

No existing laws or statutes on public sector governance and decision-making processes in the UK explicitly mention emerging AI concepts like foundation models.

No current UK laws explicitly discuss the cutting-edge AI capabilities of foundation models, general purpose AI, generative AI, large language models, artificial general intelligence or frontier models.

This is not to say that no legislation applies to foundation models. Cross-cutting legislation on data protection and equality governs their development and deployment, as discussed in the section above. And specific applications, for example 'automated decision-making', are covered under the UK GDPR.

However, a significant issue in developing regulation and controls for new use-cases and technologies is whether current regulators and other public undertakings have the necessary powers in their respective foundational documents to act. If they act to regulate, or otherwise control, such use-cases and technologies when it is not entirely clear that they have the power to do so, they risk challenge via judicial review for acting ultra vires ('beyond powers'). Many regulators of public-sector use of foundation models may therefore be operating with some uncertainty regarding whether they have the mandate to do so.

---

87 AI Law Consultancy commissioned research.

The Ada Lovelace Institute recommends that the Government clearly articulates how the principles outlined in 'A pro-innovation approach to AI regulation' will apply and be implemented in scenarios where there is no regulator with obvious current responsibility for doing so, including the public sector.[88]

Furthermore, we recommend that the Government considers the case for legislation equipping regulators with a common set of AI powers to put them on an even footing in addressing AI.

Implementing these recommendations would provide regulators with much greater certainty that they were operating withing the scope of their powers when regulating the use of foundation models in the UK public sector.

Guidance and other statements

Some departments and public bodies have begun issuing guidance and other publications that takes account of foundation models, or the related emerging AI terms listed above. These bodies are listed below.

Many of these organisations have been thinking about and issuing guidance on AI more generally for quite some time. This may explain why they are at the forefront in considering the significance of foundation models. However, this does not imply that such thinking is either complete or advanced. Many of these bodies have begun to issue related documents such as interim guidance or consultations only since the beginning of 2023.

---

88   Ada Lovelace Institute (n 44).

Public bodies giving some consideration to foundation models or related concepts

Bank of England	Government Office for Science
The Cabinet Office	The House of Lords Select Committee on Artificial Intelligence
Centre for Data Ethics and Innovation (CDEI)	The House of Lords, via its library
The Communications and Digital Committee of the House of Lords	Information Commissioner's Office (ICO)
Competition and Markets Authority (CMA)	Intellectual Property Office
Department for Business, Energy and Industrial Strategy	Medicines and Healthcare products Regulatory Agency (MHRA)
Department for Science and Innovation	NHS (England only)
Department of Education	Office for AI
Equality and Humans Rights Commission (EHRC)	OFCOM
Financial Conduct Authority (FCA)	Prudential Regulation Authority (PRA)
Financial Reporting Council.	

Several public bodies have undertaken no or only very limited public consideration of foundation models and their impact on their areas of interest. The AI Law Consultancy's report lists these bodies.

Some may question whether emerging AI concepts are immediately relevant for all public-sector organisations. But these technologies are likely to become increasingly important if uptake continues to grow at current rates. Therefore, any institutions which have not yet considered emerging AI systems need to proactively review how they may impact their work rather than addressing consequence after systems have been deployed.

Organisations already evaluating emerging AI can play a role in supporting others just starting this process. For example the members of the Digital Regulation Cooperation Forum – namely the Information Commissioner's Office (ICO), Competition and Markets Authority (CMA), Ofcom and the Financial Conduct Authority (FCA) – are all on the front foot in considering foundation models. They could collaborate with less prepared institutions, helping them anticipate and understand the implications of technologies like foundation models.

Specific examples of existing guidance

Different parts of central and local government, such as the CDDO, Department for Science, Innovation and Technology (DSIT), CDEI, and London Office of Technology and Innovation (LOTI), are already working on resources and initial guidance on the use of foundation models in public sector work.[89]

On 29 June 2023, the CDDO published 'Guidance to civil servants on use of generative AI'.[90] This guidance does not restrict civil servants from using publicly available generative AI tools. However, it urges caution, and awareness that outputs from these tools are susceptible to bias and misinformation and thus need to be checked and cited appropriately.[91]

It also instructs civil servants to never input sensitive or classified information, information which would prematurely reveal the government's intent, or personal data into publicly available generative AI tools such as ChatGPT and Bard. It specifically rules out the use of these tools for:

- '[writing] a paper regarding a change to an existing policy position [...] the new policy position would need to be entered into the tool first, which would contravene the point not to enter sensitive material'

- '[analysing] a data set [...] to present in a government paper', unless the data is publicly available or consent has been sought from the data owner.

Currently, the guidance focuses on how civil servants should use publicly available generative AI tools like ChatGPT and Bard; the considerations raised would not all necessarily apply to an internal LLM-based product hosted on government servers. But this first iteration of guidance on the more informal use of publicly available foundation-model-based tools is subject to review after six months.

---

89   'London Office of Technology and Innovation Roundtable on Generative AI in Local Government' (n 31).

90   Central Digital and Data Office (n 35).

91   The European Commission, as of May 2023, issued similar internal guidelines for staff on the use of online publicly available generative AI tools 'Guidelines for Staff on the Use of Online Available Generative Artificial Intelligence Tools' https://www.politico.eu/wp-content/uploads/2023/06/01/COM-AI-GUIDELINES.pdf accessed 16 August 2023.

What else can be done to govern foundation models in the UK public sector?

This section gives policymakers an overview of options to consider when developing governance regimes for public sector uses of foundation models.

Building on the principles and governance outlined in the previous section, we suggest additional mechanisms policymakers could implement to minimise risks, including:

- mandating independent audits
- ongoing monitoring and evaluation
- public engagement in foundation model governance
- small-scale pilots to test applications
- workforce training and education
- addressing technical dependencies
- investing in domestic computing infrastructure.

In 2020, the Committee on Standards in Public Life examined the application of the Nolan principles to the use of machine learning in public services.[92] It recommended updates to procurement processes, impact assessments, transparency, monitoring and evaluation, oversight mechanisms, and further training and education for staff.

The recommendations of that report remain a solid base which government should implement across its AI use, from foundation models to narrower applications of machine learning, to simple predictive analytics and automated decision-making.

Below, we explore interventions from further up the process of developing and acquiring foundation model capabilities, all the way through to post-deployment interventions.

---

92   Committee on Standards in Public Life, 'Artificial Intelligence and Public Standards: A Review by the Committee on Standards in Public Life' (2020) https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/868284/Web_Version_AI_and_Public_Standards.PDF accessed 10 March 2023."plainCitation":"Committee on Standards in Public Life, 'Artificial Intelligence and Public Standards: A Review by the Committee on Standards in Public Life' (2020

Procurement and technical lock-in

> Government has market power when procuring foundation model applications. It can use this to ensure that foundation models developed by private companies for the public sector uphold public standards.[93]

Provisions for ethical standards should be introduced early in the procurement process and explicitly incorporated into tenders and contractual agreements.

First, when procuring foundation-model-based products, the consideration should be the problem at hand. This may lead to identification of existing solutions whose deployment is easier than acquiring a new tool. This helps to avoid spending money on foundation model tools for their own sake.

Second, participants in our roundtable with government participants raised the issue of technical lock-in. This occurs when the public sector becomes excessively dependent on a specific technology, platform or service provider.[94] For example, when a department procures a foundation model reliant on a specific provider's data formats and cloud hosting. This risks ongoing dependency on one vendor. Dependency of this kind reduces the ability to easily and cheaply switch to another vendor.

Again, government should consider this risk during the procurement process. It should seek interoperable standards to allow for easier migration between different technologies and vendors. This would maintain flexibility, competitiveness and ethical integrity, in turn protecting the public sector from excessive costs and inability to access future innovations.

Roundtable participants were keen for the Frontier AI Taskforce to play a coordinating role across the public sector, by:

- building public sector capacity to coordinate interactions with large technology firms such as Microsoft and Google
- providing expertise and working with the Crown Commercial Service
- helping to set purchasing guidelines and providing technical expertise on standards and technical lock-in concerns.

---

93   ibid 8.
94   Ada Lovelace Institute government roundtable on the use of foundation models in the public sector (n 45).

Investing in domestic compute and data centres

Roundtable participants from government raised the continuing need for domestic compute and data centres to operate LLMs and future foundation models.[95]

They highlighted that many use cases will be limited by data privacy requirements and the confidentiality of government documents. Others have highlighted that 'Cloud [computing] located in foreign countries and owned by private foreign companies is also a source of dependence and vulnerability for the UK Government.' Risks include damage to undersea cables, price hikes, lack of oversight and control of data (especially confidential documents).[96]

For some use cases, government could seek to ensure that cloud providers locate more data centres physically in the UK. This would help account for laws and regulations that restrict the transfer of sensitive data outside the UK. Overreliance on the 'Big Three' cloud computing providers (Microsoft Azure, Amazon Web Services, and Google Cloud Platform) is a concern.

The Government has already committed to invest £900 million in wider access to computational resources for UK AI researchers, allowing them to undertake more complex modelling and exploratory research.[97]

Alternative models may need to be considered so that use of foundation models in sensitive public services like health, education and national security are well managed and accessible to academics and NGOs as well as government, making meaningful transparency and accountability central.

Mandating third-party audits before government uses of foundation models are deployed

A possible oversight mechanism is mandating independent third-party audits for all foundation model systems used in government, whether developed in-house or procured from external providers. Audits would properly scrutinise AI systems. This would help minimise risks and ensure robustness of use.

The audit requirement would apply to any firm seeking to provide foundation models to government, including providers of APIs as part of integrated products or as custom-built tools.

---

95   ibid.

96   Haydn Belfield (n 41).

97   Zoubin Ghahramani and others, 'Independent Review of The Future of Compute: Final Report and Recommendations' (2023) https://www.gov.uk/government/publications/future-of-compute-review/the-future-of-compute-report-of-the-review-of-independent-panel-of-experts accessed 16 June 2023. 'Spring Budget 2023' (2023) https://www.gov.uk/government/publications/spring-budget-2023/spring-budget-2023-html accessed 30 June 2023.

It would also apply to tools and services developed internally by central government, local government and other public bodies.

Third-party audits could serve several important functions.

First, audits allow anyone deploying foundation models and their applications to both 'anticipate harms [...] before a system is in use and [...] monitor or reassess those risks as the system changes and develops.' This allows risks to be mitigated ahead of deployment. Where risks are too great, audits can also demonstrate that the government should not continue to pursue a given use-case.[98]

Second, audits would help assess whether foundation models were appropriate for government use by their own standards. Governance, model and application audits would not necessarily directly assess a foundation model application's compliance with the Nolan principles but would offer significant evidence to enable a decision about whether this is the case.

Third, making audits a requirement to access government contracts would create a strong financial incentive for firms to comply. This would stimulate sustained and reliable demand for auditing services.[99]

Finally, it would pave the way for regulation requiring statutory audits for foundation models in the future. The audits would offer an evidence base for developing standards and best practices for this. This could help set a benchmark for the wider AI industry and create case studies of foundation model auditing methods in practice.[100]

Standards, clear examples of best practice and a growing ecosystem of auditors capable of auditing foundation models would reduce the burden on businesses to undertake their own audits. This would smooth any transition to mandatory statutory audits of foundation models, such as in high-risk domains like critical infrastructure or recruitment.

The Oxford Internet Institute and the Centre for Governance of AI have recently proposed an approach to a third-party audit which we welcome. The three-layered approach requires governance, model and application audits, which would inform and complement each other.[101] The researchers summarise these three layers as follows:[102]

---

98   Ada Lovelace Institute, AI Assurance? Methods for assessing risks, outcomes and impacts (2023)
      https://www.adalovelaceinstitute.org/report/risks-ai-systems/ accessed 16 August 2023./

99   ibid 'Enabling an ecosystem of risk assessment'.

100 ibid.

101 Jakob Mökander and others, 'Auditing Large Language Models: A Three-Layered Approach' (arXiv, 16 February 2023) 16
      http://arxiv.org/abs/2302.08500 accessed 20 June 2023.

102 ibid 19–20.

1. Governance audits: 'technology providers' accountability structures and quality management systems are evaluated for robustness, completeness, and adequacy.'

2. Model audits: 'the foundation model's capabilities and limitations are assessed along several dimensions, including performance, robustness, information security, and truthfulness.'

3. Application audits: products and services built on top of foundation models are first assessed for legal compliance and then evaluated based on impact on users, groups and the environment.

Figure 3: A three-layered approach to audits[103]

Governance audit

- Model limitations
- Model characteristics

- Training datasets
- Model selection and testing procedures

- Impact reports
- Failure model analysis

- Modal access
- Intended/prohibited use cases

Model audit

- Model limitations
- Model characteristics

Application audit

- Output logs
- Environmental data
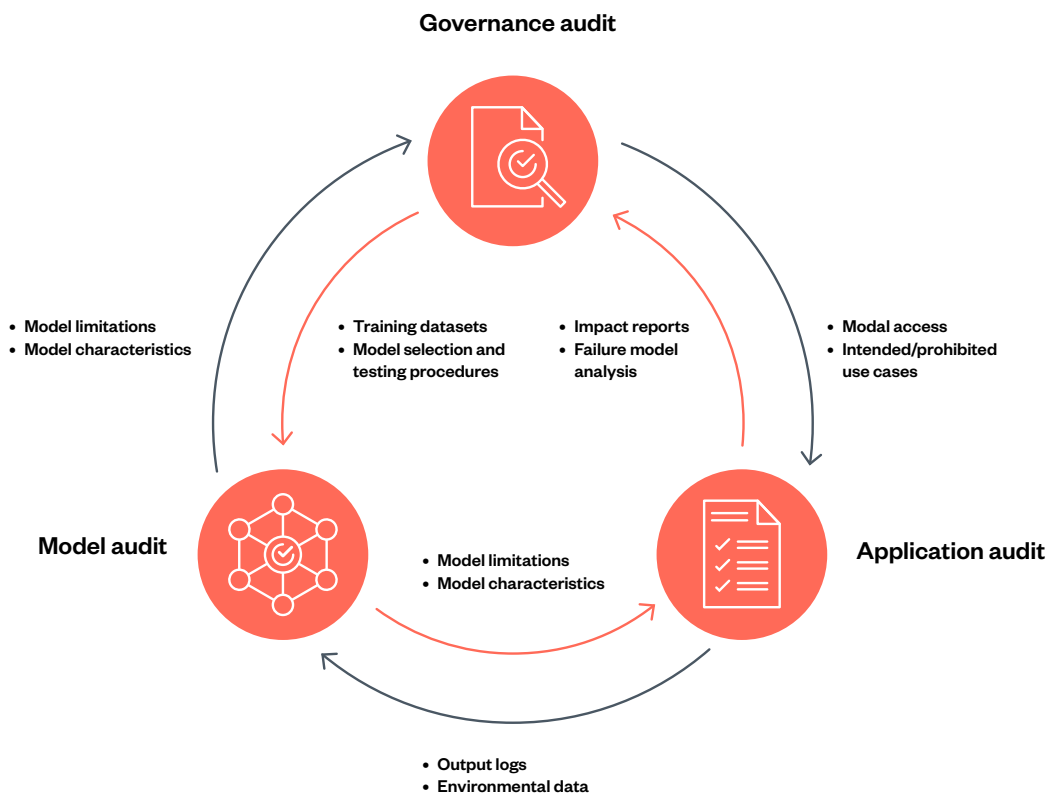
The auditing process could focus on aspects such as system robustness, identifying domain-specific limitations and potential exploits through red-teaming.

This could foster trust in the systems being used, as well as provide evidence that effective auditing of these systems is achievable.

---

103 Ibid 17

Government would need to set out standards and requirements for auditors. It could provide either certification of auditors or case-by-case approvals of audits.

This is a fast-moving area and as yet there are no accepted standards for auditing the capabilities and risks of foundational models. In the meantime, government should examine existing proposals for AI auditing, such as the ICO's AI audit framework.[104] It may be beneficial to establish an audit oversight board to monitor auditors. This board could determine the expected scope and standards for audit work.[105] Ultimately, standards are necessary – otherwise, the auditor might take on more of an advisory role for the company they are auditing, rather than serving the interests of government and the public.

The audit would ideally be published openly when a system is deployed and integrated into Algorithmic Transparency Recording Standard under Tier 2 (2.5 Risks, Mitigations and Impact Assessments) alongside any impact assessments.[106]

While the approach proposed by the Oxford Internet Institute and the Centre for Governance of AI offers advantages, there are some counterarguments.

First, even well-conducted, independent audits cannot guarantee that all risks have been identified and mitigated. Model evaluations (and therefore audits) rely on the construct validity of the metrics, such as robustness or truthfulness, used to assess the model.[107] It is a challenge to operationalise these normative concepts, let alone more complex objectives like 'the public good'.[108]

Second, audits cannot resolve inherently political choices about what trade-offs to make in implementing these systems in public services. For example, an audit could reveal if a new welfare fraud identification system reduced false positives but shifted the demographic characteristics of people incorrectly flagged by the system. But it could not determine what should be done based on that information, for example in terms of the consequences for people incorrectly flagged. These decisions are better resolved through deliberative public engagement, as discussed below.

---

104 'Guidance on AI and Data Protection' (Information Commissioner's Office, 13 June 2023) https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/ accessed 23 August 2023.

105 Inioluwa Deborah Raji and others, 'Outsider Oversight: Designing a Third Party Audit Ecosystem for AI Governance', Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (ACM 2022) 7,9 https://dl.acm.org/doi/10.1145/3514094.3534181 accessed 17 March 2023.

106 'Algorithmic Transparency Recording Standard v2.1' https://www.gov.uk/government/publications/algorithmic-transparency-template/algorithmic-transparency-template accessed 16 August 2023.

107 Jakob Mökander and others (n 105) 16.

108 Ada Lovelace Institute, Inform, Educate, Entertain... and Recommend? Exploring the use and ethics of recommendation systems in public service media (2022) https://www.adalovelaceinstitute.org/wp-content/uploads/2022/11/Ada-Lovelace-Institute-Inform-Educate-Entertain...and-Recommend-Nov-2022.pdf accessed 16 August 2023.

Third, mandatory third-party audits could increase costs and extend timelines for companies. However, audits can also add value by reducing the risks of accidents and misuse, thus reducing potential reputational risks to companies and saving them from greater future losses.[109]

Fourth, there are concerns about the quality control measures to ensure the validity of audits. Government should set out certification requirements for auditors. Over time, government should also create a standardised auditing process, with robust tools and procedures.[110]

Finally, there may be concerns about slowing down the pace of government adoption of foundation-model-based tools. Delay can be desirable if it allows for properly addressing and preparing for risks and challenges. On the other hand, auditing requirements could only be applied to systems in live deployment, when they are regularly accessed by significant numbers of government staff or the public. The auditing requirements could be waived for proof-of-concept systems or limited pilots that have their own robust risk assessment processes and a clearly defined scope and timeline, until those systems move into live deployment.

Addressing these concerns up front and developing adequate responses is an essential step in implementing this policy proposal.

Public participation

Roundtable participants raised the importance of involving the public in decision-making.[111]

> The government should incorporate meaningful public engagement into the governance of foundation models, particularly in public-facing applications. While public sector institutions have existing mandates, deploying AI systems raises new questions of benefits, risks and appropriate use that need to be informed by public perspectives.

---

109   Jakob Mökander and Luciano Floridi, 'Operationalising AI Governance through Ethics-Based Auditing: An Industry Case Study' (2023) 3 Ai and Ethics 451.

110   Inioluwa Deborah Raji and others (n 108) 7–9.

111   Ada Lovelace Institute industry and civil society Roundtable on the use of foundation models in the public sector.

Past research by the Ada Lovelace Institute found that public participation can foster accountability and broaden impact considerations. But it remains unclear how corporate AI labs and public sector bodies can implement different kinds of public participation methods in the design, deployment and governance of these systems.[112]

In recent research, we found that commercial AI labs do not regularly or consistently do this, and that these practitioners lack a common set of methods. Using public participation methods is even more challenging in the context of foundation models that lack a clear context of use, making it harder for the public to confidently assess potential impacts and social needs.[113]

Government and civil society organisations can play a role in collaboratively establishing standards for participation, built on successes in contexts like algorithmic impact assessments in healthcare. This would strengthen government's use of foundation models and contribute best practices that industry could emulate, to avoid industry unilaterally defining terms and approaches for participation.

Citizens' assemblies and participatory impact assessments focused on foundation models could allow joint deliberation on proposed uses and make recommendations. Authentic involvement provides transparency and democratic legitimacy. It steers deployment based on societal needs and values, not just government or corporate incentives.

Decisions about developing and using foundation models ultimately involve political choices, including:

- whether to spend the considerable sums of money required create and use foundation model and applications
- whether to increasingly delegate judgements about what information to prioritise in government decision-making.
- what information should foundation model applications share with members of the public
- whether to use foundation model capabilities to create increasingly targeted, means-tested, proactive public services.

Public participation can help government make informed, considered decisions about these trade-offs.

112   Ada Lovelace Institute, Algorithmic Impact Assessment: A case study in healthcare' (2022) https://www.adalovelaceinstitute.org/report/algorithmic-impact-assessment-case-study-healthcare/ accessed 19 April 2022. Ada Lovelace Institute, The Citizens' Biometrics Council (2021) https://www.adalovelaceinstitute.org/report/citizens-biometrics-council/.

113   Lara Groves and others, 'Going Public: The Role of Public Participation Approaches in Commercial AI Labs', 2023 ACM Conference on Fairness, Accountability, and Transparency (ACM 2023) https://dl.acm.org/doi/10.1145/3593013.3594071 accessed 16 August 2023.

Pilots

Roundtable participants from industry, civil society and academia suggested government should take a cautious approach to implementing foundation models.

> Government departments should start with less risky, discrete pilot use cases. These use cases should be easily understandable even by relatively technically unsophisticated users.[114]

This could include trialling chatbots to help service users find relevant policies and publications by searching and retrieving publicly available government documents.

As government becomes more proficient with the technology, it could progress to conducting large-scale pilots and A/B testing. This would be especially helpful for higher-risk use cases such as auto-generating emails to communicate with residents, where the potential benefits and harms are higher.[115]

Local authority representatives were keenly interested in having a shared repository of use cases. This could be hosted by the Frontier AI Taskforce, possibly in collaboration with the Local Government Association or Socitm.[116] This could prevent unnecessary duplication of effort, saving time and financial resources. It should include regional perspectives, to ensure relevant and localised insights.

Finally, participants suggested that a shared collaborative environment for executing individual experiments could be beneficial, allowing local authorities or smaller public bodies to pool resources and contribute to and learn from various test cases, again reducing duplication of effort in piloting the use of foundation models.[117]

Training and education

Some roundtable participants noted the difficulty of recruiting skilled developers and the challenge of offering competitive salaries in the public sector, despite recent redundancies in big technology companies.

---

114   Ada Lovelace Institute industry and civil society roundtable on the use of foundation models in the public sector.

115   'Ada Lovelace Institute government roundtable on the use of foundation models in the public sector.

116   London Office of Technology and Innovation roundtable on generative AI in local government.

117   ibid.

Others highlighted that, even when the data science and product teams in government understand the details of foundation models, officers (in legal, commercial, data protection or compliance) and senior management responsible for overseeing their use often lack the knowledge and understanding to fully assess their application.

Public- and private-sector providers of public services should ensure that employees developing, overseeing or using foundation models systems undergo continuous training and education,[118] so that they can proficiently engage with suppliers, provide appropriate risk management and compliance support, and ensure that foundation model applications serve the public as well as possible.

The government has recognised and taken steps to address civil service skills deficits, especially relating to digital, data and technology capabilities. For example, the Digital Secondment Programme plans to bring in outside experts from large technology companies.[119] This secondment programme could target foundation model experts who could work directly on public service applications and share their expertise with permanent civil service teams, when the Frontier AI Taskforce is wound up and no longer serving that purpose.

Continuous monitoring and evaluation

> Monitoring and evaluation needs to continue beyond initial development and procurement of foundation model applications.

It can take a similar form to the auditing process outlined above but would include recording real-life data and feedback from the operation of the applications.

Local and central government representatives in our roundtables felt that this continuous monitoring and evaluation by public services, of both public and private applications, was needed to ensure foundation model systems always operate as intended.[120]

---

118   Committee on Standards in Public Life (n 96) 9.
119   'Whitehall Set to Bring in AI and Data Experts under Plans to Turbocharge Productivity' (GOV.UK)
        https://www.gov.uk/government/news/whitehall-set-to-bring-in-ai-and-data-experts-under-plans-to-turbocharge-productivity
        accessed 21 July 2023.
120   London Office of Technology and Innovation roundtable on generative AI in local government. See also similar recommendations
        made by in: Committee on Standards in Public Life (n 96) 9.

Transparency

Research by the Ada Lovelace Institute and The Alan Turing Institute,[121] and by the CDEI,[122] shows the public have clear expectations about the transparency of AI systems, and that this is crucial to whether these technologies are perceived as trustworthy.

Rolling out the Algorithmic Transparency Recording Standard[123] across the public sector would ensure that the government's use of these models lives up to public expectations. It would also support monitoring and evaluation by allowing better external, independent scrutiny. It would provide a more systematic understanding within and outside government of how widespread foundation model applications are across government.

Regular reviews of legislation, regulation and guidance

Governance mechanisms themselves, including guidance, need to be reviewed regularly. The AI Law Consultancy's survey identified a pressing need to regularly review public bodies' work to see how they are keeping up to date with the implications of foundation models and future developments in emerging AI.[124] It suggests public sector bodies review their policies and progress at least twice yearly, given the substantial potential impacts of AI on rights and ethics. The outcomes of the reviews should be published, to promote transparency and accountability.

Some legislative changes could also help empower regulators to act appropriately on emerging forms of AI within their remits. As discussed above, uncertainties around powers create a risk of judicial review. Solutions include expanding regulators' statutory functions or amending legislation like the Legislative and Regulatory Reform Act 2006.

In other research, Ada has made recommendations on ensuring legislation keeps up with the development of foundation models in the public sector.[125] These include:

- Review the rights and protections provided by existing legislation such as the UK General Data Protection Regulation (GDPR) and the Equality Act 2010 and – where necessary – legislate to

---

121   Ada Lovelace Institute and The Alan Turing Institute, How Do People Feel about AI? (2023) 10 https://www.adalovelaceinstitute.org/report/public-attitudes-ai/ accessed 16 August 2023.

122   BritainThinks and Centre for Data Ethics and Innovation, 'AI Governance' (2022) https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1146010/CDEI_AI_White_Paper_Final_report.pdf.

123   CDDO and CDEI, 'Algorithmic Transparency Recording Standard - Guidance for Public Sector Bodies' (GOV.UK, 5 January 2023) https://www.gov.uk/government/publications/guidance-for-organisations-using-the-algorithmic-transparency-recording-standard/algorithmic-transparency-recording-standard-guidance-for-public-sector-bodies accessed 9 February 2023.

124   AI Law Consultany commissioned research.

125   Ada Lovelace Institute (n 44).

introduce new rights and protections for people and groups affected by AI to ensure people can achieve adequate redress.

- Introduce a statutory duty for legislators to have regard to the principles, including strict transparency and accountability obligations.

- Explore the introduction of a common set of powers for regulators and ex ante, developer-focused regulatory capability.

- Clarify the law around AI liability, to ensure that legal and financial liability for AI risk is distributed proportionately along AI value chains.

- Allocate significant resource and future parliamentary time to enable a robust, legislatively supported approach to foundation model governance as soon as possible.

- Review opportunities for and barriers to the enforcement of existing laws – particularly the UK GDPR and the intellectual property (IP) regime – in relation to foundation models and applications built on top of them.

In summary, regular ongoing review is needed of emerging AI systems and governance mechanisms. Public reporting would motivate regulators to keep pace with technology. Targeted legal changes could strengthen their ability to oversee emerging AI capabilities.

How might foundation models evolve in the future and what are the potential implications?

The capabilities of foundation models are rapidly evolving. It is important to consider how foreseeable technological developments could impact their application and governance in the public sector. Maintaining responsible oversight requires governance regimes flexible enough to ensure public sector uses uphold ethical principles as capabilities progress.

This section aims to help policymakers and public sector leaders anticipate and respond to the shifting foundation model landscape. It presents roundtable participants' perspectives on the long-term future of foundation models and their use in government, not just the immediate opportunities and risks.

We then discuss potential trends in foundation model evolution and their implications, based on a review of recent conference papers, demos and live products. This is not a comprehensive attempt at foresight but is meant to broaden the horizons of those deploying public sector foundation models beyond the next 6–12 months.

First, we examine how tool-assisted foundation models with access to external data sources and text-to-action models with increased autonomy may alter use cases and introduce new risks. Second, we analyse different potential trajectories for the foundation model ecosystem, towards either concentration in a few private companies or a proliferation of open source alternatives, and how this could shape public sector procurement and policy.

Roundtable views on the future of foundation models and their applications

We asked roundtable participants to consider the long-term future of foundation models, not just immediate opportunities and risks. Responses suggested limited thinking had been given to horizon scanning or discussion of potential unintended long-term consequences of current and near-term foundation model use. Government needs to give more thought to the longer-term development of

these technologies, beyond the next year.[126]

Participants raised questions and concerns about how foundation models may evolve over time; for instance, whether they will become as ubiquitous as social media, follow the hype cycle of technologies like blockchain that have failed to broadly materialise, or become more specialised tools that are highly disruptive for certain activities but not others.

The speed of change means governance mechanisms often lag behind cutting-edge capabilities. Participants worried that local and central government leadership lacked resources and incentives to continuously monitor and understand emergent technologies. They emphasised the need for continuous training and upskilling, so that oversight roles like legal, policy and ethics specialists understand rapid changes.

There was uncertainty around future business models and whether there would be a few large foundation model providers or a less concentrated market for foundation models, better data portability across government, and methods for codifying laws and rules in a machine-readable format.

As well as addressing the current impact of novel technologies is important, having an eye to the future is important, particularly when it is arriving faster than ever before. With that in mind, we undertook desk research and found some areas policymakers could monitor. These include developments like tool-assisted foundation models, which can interact with external systems like web browsers or databases, and text-to-action systems, which convert instructions into automated behaviours. These developments may raise additional concerns around accountability and control.

Below, we discuss what policymakers and deployers of foundation models in government might need to think about when looking at the future of these systems. This is from a technical capabilities perspective and from the perspective of the broader market and ecosystem of foundation models.

Tool-assisted foundation models and text-to-action models

One potential future of foundation models may be an increased capability to interact with external tools and websites, from search engines to calculators, maps and even other specialised AI

---

126   Ibid. In our 'Regulating AI in the UK' report, we argue that Government is currently largely reliant on external expertise to understand developments in AI. We argue that Government understanding of the sector, and of necessary governance interventions, would be strengthened by conducting systematic in-house analysis. We propose that the Frontier AI Taskforce should invest immediately in small pilot projects that could begin to build this in-house expertise and infrastructure and which – if successful – could be continued as part of the central functions. See also: Ada Lovelace Institute, Keeping an eye on AI: Approaches to government monitoring of the AI landscape (2023) https://www.adalovelaceinstitute.org/report/keeping-an-eye-on-ai/ accessed 23 August 2023.

systems.[127] A current example is OpenAI's release of ChatGPT Plugins, which enables a user to input a text prompt to ChatGPT to execute commands with specific websites, such as buying food from shopping sites like Shopify or booking travel though sites like Kayak.com.[128]

Researchers are starting to explore what capabilities could be enabled by giving foundation models the ability to interact with external tools and plugins.[129]

Some theoretical work suggests that these abilities may reduce the risk of the 'hallucinations' discussed earlier. Access to domain-specific tools could also enable foundation models to do a wider range of more niche tasks.

Some research considers whether access to external tools could improve interpretability and robustness. For example, end users could see how and when different tools are used, and how those tools contribute to the final output of the foundation model. This could make the process of creating the output more transparent.

External tools may also enhance the robustness of foundation models' outputs if they allow systems to adapt in response to unpredictable real-world environments. In environments where there is domain shift, away from the distribution in the foundation model's original training dataset, the model could access new data relevant to the changed environment. This could happen, for example, if a model trained only on pre-pandemic data was able to use search engines to access data about changing habits after the onset of the pandemic.

These examples are hypothetical. But given the speed at which these models develop, government should pre-empt potential new capabilities rather than waiting for them to appear.

A related example is text-to-task systems, in which natural language instructions are turned, through an API, into actions on a screen or physical actions using a robot.[130] This could substantially increase the number and variety of tasks that foundation models can be used for. This would also increase the gap between human users and their actions, reducing people's autonomy and control. Oversight, accountability and explainability would therefore be even more important.

While the potential of these models may prompt enthusiasm, caution is needed. A full risk assessment would be needed before applications are even piloted for public services. Rigorous testing would be needed, and a clear understanding of how legal accountability and redress would

---

127 Yujia Qin and others, 'Tool Learning with Foundation Models' https://arxiv.org/abs/2304.08354 accessed 27 June 2023."

128 'ChatGPT Plugins' (OpenAI, 23 March 2023) https://openai.com/blog/chatgpt-plugins accessed 21 August 2023.

129 Yujia Qin and others (n 127)

130 Adept, 'ACT-1: Transformer for Actions' (14 September 2022) https://www.adept.ai/blog/act-1 accessed 10 July 2023; Anthony Brohan and others, 'RT-1: Robotics Transformer for Real-World Control at Scale' (arXiv, 13 December 2022) http://arxiv.org/abs/2212.06817 accessed 10 July 2023.

function (if at all) in any applications that use these kinds of delegated action systems. Government must keep an eye on the future and routinely complete horizon scanning and futures thinking exercises to consider both risks and opportunities.

Will the ecosystem of foundation models move towards more powerful open-source models or an oligopoly of leading labs for the most capable systems?

Possible futures for the market structure of foundation model development range from large private labs continuing to dominate provision, with an oligopoly of a few big models used across society, to a proliferation of open-source foundation models, each fine-tuned to a different domain or use case.[131]

Researchers at the Brookings Institution in Washington DC, have found that the market for cutting-edge foundation models strongly tends towards market concentration. The fixed costs of training a foundation model are high and the marginal costs of deployment are comparatively low, creating large economies of scale. This favours a few large companies dominating the market.[132]

Specifically, Epoch AI estimates that the compute used to train state-of-the-art machine learning models increased by about eight orders of magnitude (that is, 100 million times over) between 2012 and 2023.[133]

---

131   The terms 'open source' and 'open' as applied to AI systems are used divergently, often blending concepts from open source software and open science. There is currently no agreed definition of what constitutes 'open' AI, despite growing attention to the concept. See: David Gray Widder, Sarah West and Meredith Whittaker, 'Open (For Business): Big Tech, Concentrated Power, and the Political Economy of Open AI' (17 August 2023) 3–5 https://papers.ssrn.com/abstract=4543807 accessed 29 August 2023.

132   Jai Vipra and Anton Korinek, 'Market Concentration Implications of Foundation Models: The Invisible Hand of ChatGPT' (September 2023) 9 https://www.brookings.edu/wp-content/uploads/2023/09/Market-concentration-implications-of-foundation-models-FINAL-1.pdf accessed 9 July 2023.

133   Jaime Sevilla and others, 'Compute Trends Across Three Eras of Machine Learning' (arXiv, 9 March 2022) http://arxiv.org/abs/2202.05924> accessed 16 August 2023.9 March 2022

Figure 4: Compute used to train machine learning models[134]
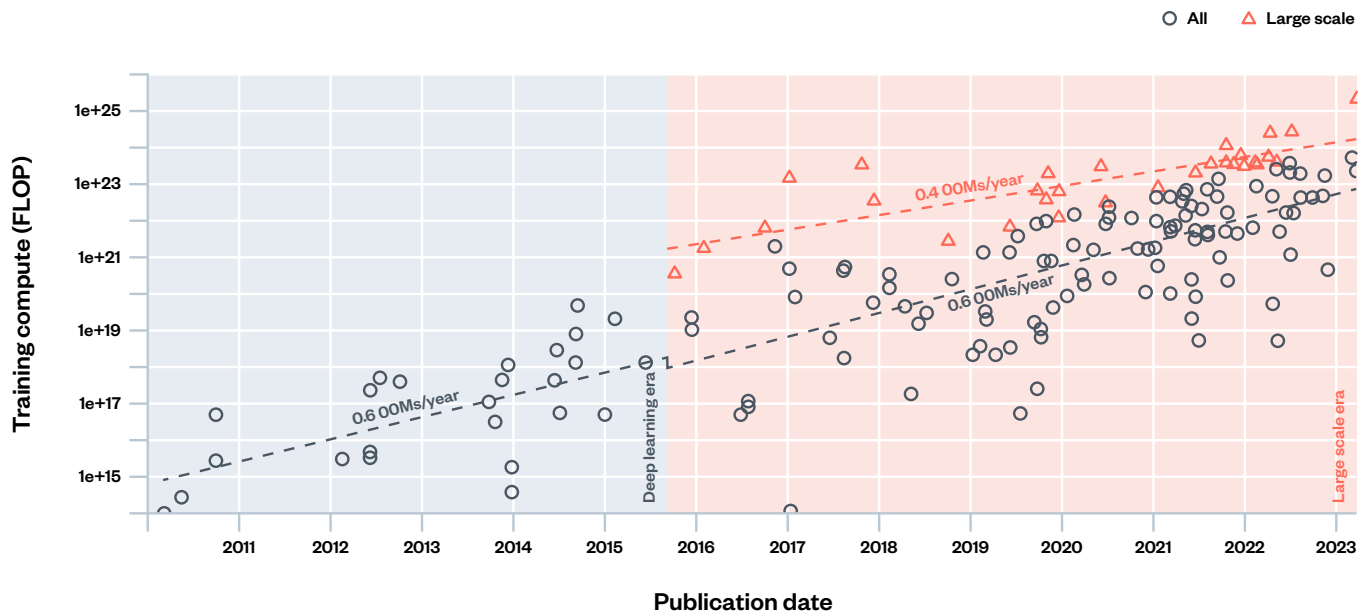
Figure 5: Estimated training compute cost in US Dollars[135]
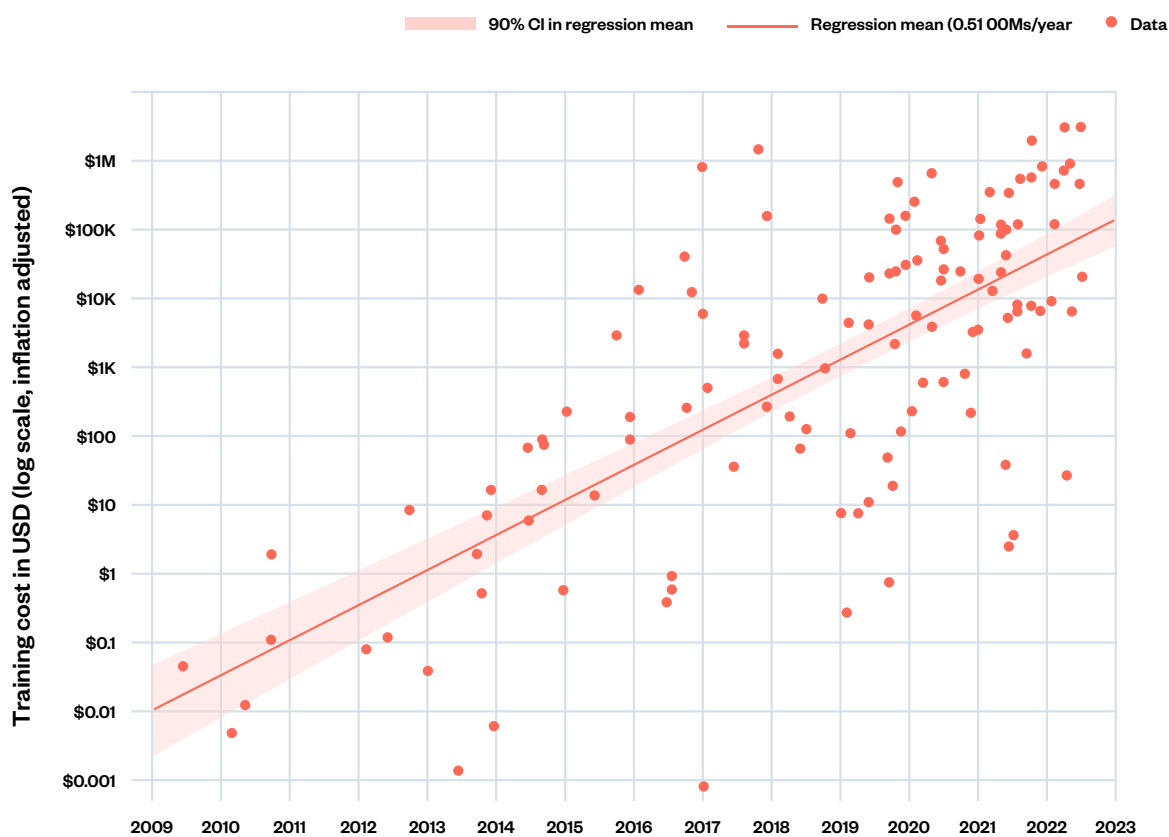
134 'ML Input Trends Visualization' (Epoch) https://epochai.org/mlinputs/visualization

135 Cottier B, 'Trends in the Dollar Training Cost of Machine Learning Systems' (Epoch, 31 January 2023) https://epochai.org/blog/trends-in-the-dollar-training-cost-of-machine-learning-systems

If the amount and associated cost of compute to train the next generation of models continues to increase, this would favour large technology firms (and potentially governments) that can afford massive investments in compute infrastructure and talent.

Researchers at Brookings predict there will continue to be a concentrated market for the most advanced foundation models, and more competition for smaller, less advanced models.[136]

Others suggest that potential physical constraints on hardware may reduce this concentration at the frontier. Limitations in transistor density and memory bandwidth may curtail growth in model scale and compute needs this decade. Sarah Constantin, Director of Corporate Development at Nanotronics, has argued that, due to a combination of limitations in transistor density and memory bandwidth, it is possible that the compute used in cutting edge models today, could grow as little as three times by 2030 and 17 times by 2040 – or less than one more order of magnitude growth this decade.[137]

Instead, progress may come from software and model architecture innovations enabling more efficient model architectures. This could allow a wider range of organisations to develop capable foundation models.

Currently, the high cost involved in both training and running inferences on foundation AI models at scale presents a significant barrier to reusability. This is true for even the most maximally 'open' AI systems.[138] However, the open-source community has already made some advances in compact, economical language models.

If LLMs do not have to be large to be effective and economical to use, then some believe that the development of LLMs need not remain in the hands of a few of large technology companies. Future progress will not necessarily come from greater buying power or access to compute and data, but instead from algorithmic improvements, and so foundation model development may decentralise.[139]

Intervention by competition regulators like the UK's Competition and Markets Authority and the USA's Federal Trade Commission could also contribute to a more fragmented market and a greater role for open-source foundation models.[140]

---

136   Jai Vipra and Anton Korinek (n 135) 19.

137   Sarah Constantin, 'The Transistor Cliff' [2023] *Asterisk* https://asteriskmag.com/issues/03/the-transistor-cliff accessed 1 August 2023.

138   David Gray Widder, Sarah West and Meredith Whittaker (n 134) 7.

139   Sarah Constantin (n 138).

140   Competition and Markets Authority, (2023), *AI Foundation Models: initial review*, https://www.gov.uk/cma-cases/ai-foundation-models-initial-review (Accessed: 1 August 2023) Staff in the Bureau of Competition & Office of Technology, (2023), *Generative AI Raises Competition Concerns*, https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/06/generative-ai-raises-competition-concerns accessed: 1 August 2023

While open source can spur innovation, it has been noted by academics from Carnegie Mellon, AI Now and the Signal Foundation that history shows that it often entrenches large technology companies' dominance.[141] Powerful technology companies can promote openness to shape policy, claiming it will make AI more democratic. But they can also exploit openness to set technical standards, integrating innovations into their proprietary products. They also benefit enormously from free open-source labour.

So even maximally open AI does not guarantee fair access or competition. Openness alone cannot address issues of oversight or industry concentration. Some argue that open source risks reinforcing existing leaders' advantages unless coupled with policies on equitable access to resources.[142]

Government should ensure its procurement and foundation model infrastructure development and integration is robust in a future dominated by a few large companies, a more decentralised foundation model ecosystem, or anything in between.

---

141   David Gray Widder, Sarah West and Meredith Whittaker (n 134) 13.
142   ibid 19.

Conclusion

Foundation models are a significant evolution in AI capabilities and could transform public services and improve outcomes for people and society. But their deployment in central and local government involves risks and challenges that must be carefully considered.

Enthusiasm about using foundation models in automated document analysis, supporting decision-making, improving management of public enquiries and knowledge management needs to be balanced against the risks of bias, privacy, security, environmental impact and workforce displacement.

This evidence review has aimed to provide an overview of the current state of foundation models in government and highlight the most pressing issues for policymakers. Priorities include the need to:

- Establish clear principles, guidance and impact assessments for public sector uses of foundation models. Existing data ethics frameworks are a baseline but may need to be enhanced to address the specific properties of foundation models.

- Start with small-scale pilots and experiments, focusing on lower-risk use cases, and learn lessons from this before any wider-scale deployment.

- Mandate independent audits of systems before they go live and throughout their lifecycle, to validate their safety, security and social impacts. This could stimulate the growth of the UK AI auditing industry.

- Monitor systems continually through red teaming, user feedback channels and evaluating real-world performance data.

- Involve the public in governance and take note of their perspectives on the benefits and risks, particularly for public-facing applications.

- Invest in skills, education and awareness at all levels of government to ensure AI is used responsibly.

Foundation models may offer an opportunity to address certain challenges in public services delivery. But government must take coordinated action to develop and deploy them responsibly, safely and ethically. This review represents a start at mapping out the issues these systems raise in public sector contexts, but further research and policy development is still required in this fast-evolving field.

Methodology

Research questions

This evidence review aimed to explore the opportunities, risks and governance required in using foundation models in the public sector. We set out to answer the following questions:

1.   What is a foundation model, and how do we define the terminology around this technology?

2.   How are foundation models already being deployed by government, formally and informally? What uses are in development or being considered?

3.   What problems/opportunities in public services delivery do local and national government see foundation models solving? Could these problems be tackled using foundation models or are better, less unstable and better regulated tools already available?

4.   What are the risks for governments and their autonomy, the unintended consequences and the limitations of foundation models?

5.   How should government use of these technologies be governed? What hard rules or guidance should governments follow?

Methods

We set out to answer these research questions as follows:

1.   Desk-based research looking at literature, news reports, transparency documents etc. outlining current or near-term government use of foundation models. This primarily addressed research question 2 but also informed the answer to questions 3 and 4.

2.   Desk-based research on definitions, opportunities, risks and use cases of foundation models in general, with analysis to apply that to government/public sector context. This addressed research questions 1, 3 and 4 and 5, to a degree.

3.   Two expert roundtables to gain input from: government sources, with greater insight into government plans to use these systems but also potentially unwilling to go beyond agreed lines; and from external experts, who are likely more open and independent of government

but may have their own preconceptions of the appropriateness of these systems. The first roundtable addressed research questions 2, 3 and 4, the second primarily question 4 and potentially 5.

4.   We commissioned the AI Law Consultancy to do a desk-based review of UK legislation, regulations and guidance, identifying where specific concepts (listed below) were directly addressed or mentioned. This provided evidence in answering research question 5.

The research did not start with a specific hypothesis. It aimed initially to give a descriptive account of current information on government use of foundation models, supplemented with (again primarily descriptive) analysis of use cases, opportunities and risks, and mentions in current legislation, regulation and guidance. We concluded with provisional recommendations to the UK government.

Acknowledgements

AI Law Consultancy

We would especially like to thank Robin Allen and Dee Masters of the AI Law Consultancy, and Ameer Ismail at Cloisters Chambers, for their supporting work analysing how foundation models and related AI concepts are addressed by the UK's legislation, regulation and guidance. This work is an important undertaking in its own right, highlighting the gaps in foundation model governance and the need for further regulation.

Appendix 1: Examples of potential public sector applications of foundation models

The wide-ranging capabilities of foundation models in text and image generation mean that they can be applied to a number of tasks, including:

providing a first draft or writing assistance when producing emails, briefings, reports etc.

- summarising
- translation
- search and retrieval
- synthetic data generation.

Applications

These capabilities can be applied to a particular domain, with the foundation model performing one or more of those tasks, either through a stand-alone interface or as part of a wider product or service. The foundation model can then be applied to numerous use cases.

Below, we outline proposed or potential use case areas and examples, drawing on conservations with people across central and local government, and public bodies and use cases suggested by think-tanks and industry.[143]

---

143 Analysis and Research Team (n 79) 9–10. Benedict Macon-Cooney and others, 'A New National Purpose: AI Promises a World-Leading Future of Britain' (Tony Blair Institute for Global Change 2023) https://www.institute.global/insights/politics-and-governance/new-national-purpose-ai-promises-world-leading-future-of-britain accessed 13 June 2023; London Office of Technology and Innovation roundtable on generative AI in local government; Centre for Data Ethics and Innovation roundtable on use of foundation models in the public sector; Ada Lovelace Institute government roundtable on the use of foundation models in the public sector; Ada Lovelace Institute industry and civil society roundtable on the use of foundation models in the public sector; Matthew Upson, 'ChatGPT for GOV.UK' (MantisNLP, 15 March 2023) https://medium.com/mantisnlp/chatgpt-for-gov-uk-c6f232dae7d accessed 20 April 2023;"plainCitation":"Matthew Upson, 'ChatGPT for GOV.UK' (MantisNLP, 15 March 2023 Ben Wodecki, 'UK Wants to Tap Large Language Models to Reinvent Government' AI Business (8 March 2023) https://aibusiness.com/nlp/uk-wants-to-tap-large-language-models-to-reinvent-government accessed 21 August 2023.

Key areas	Examples
Government communication and public enquiries	• Powering chatbots and virtual assistants, addressing basic questions and issues from the public round-the-clock • Assisting call centre employees, for example, helping them find relevant information or providing advice on handling a particular enquiry • Creating personalised text-based communications for different demographic groups or in different languages • Recommending government services to people based on relevance • Helping to signpost callers based on sentiment analysis
Document and text analysis	• Identifying key information in complex documents such as legal contracts, to reduce review times • Summarising and categorising documents • Automating repetitive tasks such as copying and pasting information from websites • Assisting in producing documents such as reports • Transcribing audio and summarising transcripts • Improving text by ensuring consistency or removing e.g. gendered language
Data analysis	• Analysis and reporting based on financial and accounting data • Real-time data analysis about service provision, changes in take-up, impact assessments and fraud monitoring • Analysing trends in healthcare reports • Synthesising large amounts of data, such as call logs • Identifying people eligible for a public service but not accessing it
Decision support	• Assisting in decision-making by evaluating grant applications or welfare support eligibility • Triaging casework by summarising and suggesting categorisation of cases to assign them to appropriate specialists
Coding assistance	• Assisting in interpreting or refactoring legacy code in government systems
Human resources	• Screening CVs and matching candidates for recruitment • Conducting initial screening interviews for recruitment
Knowledge management, search and retrieval	• Semantic search as an interface to a Corporate Memory Bank, looking for similar previous work and preventing organisational memory loss • Creating an in-house wiki by bringing together policy information from documents across departments, news, reports etc. • Summarising, categorising and tagging internal documents
Policy explanation and drafting	• Explaining large amounts of policy quickly • Converting the style of a text into 'Whitehall speak' for those who are not familiar with government style and tone • Presenting pros and cons of different options for policy document outlines, for civil servants to choose from • An adversarial drafting tool highlighting possible flaws, omissions and counterarguments for policy paper drafts • Optimising text, e.g. by removing gendered language or ensuring a consistent reading age

Case studies

The more detailed examples below represent actual or proposed foundation model applications, with details changed to make them more generally applicable and protect confidentiality. We shared these examples with roundtable participants as thinking prompts

Case study 1: AI-Assisted government briefing tool

The use of a LLM to assist staff with summarising content and drafting briefings. The proposed system could be capable of processing documents of various types, including speeches, policy reports, media analysis etc. and producing outlines for specific outputs, such as briefings for senior staff, based on instructions. This example shows the potential of foundation models in streamlining administrative search and summarisation tasks, and first draft generation, providing support for back-end processes.

Case study 2: Synthetic data generation tool

A large language model (LLM) could create synthetic datasets of imagined people engaging with public services. The LLM could be fine-tuned on existing correspondence and case notes, and then prompted with census data to generate individuals with various demographics, conditions and personal circumstances. These datasets could then be used to assess the appropriateness, privacy and security risks associated with the usage of actual correspondence with the public. This could help establish necessary governance procedures before any personal data is actually shared externally. This case study illustrates the potential for foundation models in facilitating data generation and enhancing privacy considerations in sensitive sectors.

Case study 3: Large-scale multilingual data analysis

Foundation models with text and image input capabilities could be used to interpret large and complex datasets, which could potentially involve multiple languages and limited metadata. Using self-hosted foundation models based on open-source models would allow government users to process documents classified as Official-Sensitive or above. This use of AI demonstrates its ability to search, retrieve and translate large datasets, offering valuable assistance to internal government users attempting to analyse large, multilingual and poorly structured text and image data.

Case study 4: Queries from the public pilot project

A public body could pilot a small-scale project to automate the conversion of natural language queries from the public into structured queries for their data system. It would then present that information to the public in an accessible format, allowing the public to gain some knowledge on topics they are unsure about. This approach could be used to keep a close eye on a project's implementation, assessing cost effectiveness and accuracy for example.

Appendix 2: AI Law Consultancy commissioned report

'Next generation' artificial intelligence concepts and the public sphere

Introduction

1. This Report has been prepared by the AI Law Consultancy (the Consultancy)[144], based at Cloisters' Barristers Chambers, at the request of the Ada Lovelace Institute (the Institute). It provides a snapshot of the extent to which 'next generation' Artificial Intelligence concepts are being addressed by the United Kingdom's major public undertakings by online research conducted between 27 June and 19 July 2023.

2. The Institute is concerned to know the extent to which such new technologies - beyond basic Artificial Intelligence (defined below) - were being actively debated and considered within the public sphere in the UK. With this information, the Institute will be able to identify gaps and limitations that may expose weaknesses undermining, or showing insufficient control of, the development of such technologies. This is an issue which will both concern national governance and also the extent to which the United Kingdom can be accepted as a destination for the future use and development of these technologies.

3. It is hoped that this report can therefore contribute to the discussion about the need for, and best ways to progress, further regulatory action in the UK at a time when the UK is considering next steps following the publication of the UK Policy White Paper "A pro-innovation approach to AI regulation".[145]

---

144 Robin Allen KC and Dee Masters.

145 "A pro-innovation approach to AI regulation", Command Paper Number: 815, HMSO, ISBN: 978-1-5286-4009-1, 29 March 2023, updated 22 June 2023.

Concepts

4.   There are a range of new technologies that have recently been the subject of national discourse. The Institute has identified a group of such technologies for our consideration, to which we have referred collectively in our Report as the "Concepts".

5.   These are -

- General Purpose Artificial Intelligence (**GPAI**),
- Generative AI,
- Large Language Models (**LLMs**),
- Foundation Models,
- Artificial General Intelligence, and
- Frontier Models.

6.   There is no single universal definition for each of the Concepts,[146] however the Institute, which is well aware of the range of such developments, has adopted working definitions, which we have set out in **Box A**. We have used these definitions, which were provided on 28 June 2023, to assist with our research for this snapshot Report[147]

Box A – Concept descriptions

Artificial intelligence (AI)	Artificial intelligence is a term that describes the use of computers and digital technology to perform complex tasks commonly thought to require intelligence. Artificial intelligence systems typically analyse large amounts of data to take actions and achieve specific goals, sometimes autonomously (without human direction).
General-purpose AI	General-purpose AI models are AI models that are capable of a wide range of possible tasks and applications. They have core capabilities, which become general through their ability to undertake a range of broad tasks. These include translating and summarising text; generating a report from a series of notes; drafting emails; responding to queries and questions; and creating new text, images, audio or visual content based on prompts.

---

146   If there had been such definitions it would have assisted with the process of review by introducing a degree of standardisation.

147   The Institute published, on 17 July 2023, a document entitled "Explainer: What is a foundation model?" which contains definitions of the Concepts (and more). It is an important steps towards the creation of standardised definitions in an area which is currently fluid.

Generative AI	Generative AI refers to AI systems that can augment or create new and original content like images, videos, text, and audio. Generative AI tools have been around for decades. Some more recent generative AI applications have been built on top of general purpose (or foundation) models, such as OpenAI's DALL-E or Midjourney, which use natural language text prompts to generate images.
Large language model	Language models are type of AI system trained on massive amounts of text data that can generate natural language responses to a wide range of inputs. These systems are trained on text prediction tasks. This means that they predict the likelihood of a character, word or string, given either its preceding or surrounding context. For example, predicting the next word in a sentence given the previous paragraph of text. Large language models now generally refer to language models that have hundreds of millions (and at the cutting edge hundreds of billions) of parameters, which are pretrained using a corpus of billions of words of text and use a specific kind of Transformer model architecture.
Foundation model	The term 'foundation model' was popularised in 2021 by researchers at the Stanford Institute for Human-Centered Artificial Intelligence. They define foundation models as 'models trained on broad data (generally using self-supervision at scale) that can be adapted to a wide range of downstream tasks'. In this sense, foundation models can be seen as similar to 'general purpose AI models', and the terms are often used interchangeably.
Artificial general intelligence	Artificial general intelligence (AGI) is a contested term without an agreed definition. Researchers from Microsoft have sought to define AGI as 'systems that demonstrate broad capabilities of intelligence, including reasoning, planning, and the ability to learn from experience, and with these capabilities at or above human-level'. A stronger form of AGI has been conceptualised by some researchers as 'high-level machine intelligence', or when 'unaided machines can accomplish every task better and more cheaply than human workers'.
Frontier model	Frontier models are a category of AI models within the broader category of foundation models. There is no agreed definition of what makes a foundation model a 'frontier' model, but it is commonly thought of as the model which are the most effective at accomplishing specific distinct tasks, such as generating text or manipulating a robotic hand, often using cutting-edge or state of the art techniques.

Legal resources

7.   The first snapshot we undertook concerned the extent to which primary and secondary legislation[148] in the UK was already written in a way that specifically addressed new forms of 'next generation' AI.[149] The reason for this snapshot was, in summary, to identify the extent to which current legislation discussed the Concepts.

8.   This snapshot involved a desk top review of two standard well-recognised legal resources, Westlaw and LexisNexis. This research was conducted under our supervision by Ameer Ismail, our colleague at Cloisters. The results are summarised in Appendix A.

9.   A major issue in developing regulation and controls for new use-cases and new technologies is the extent to which current regulators and other public undertakings have the necessary powers in their respective foundational documents to act. This is important because if they act to regulate or otherwise control such new use-cases and technologies when it is not entirely clear that they have the power to do so, they risk being challenged by judicial review on the basis that they have acted *"ultra vires"* (literally meaning "beyond powers").

10.   While the four Digital Regulators that comprise the Digital Regulation Cooperation Forum[150] (DRCF) have significant specific statutory powers to address data use–cases when these involve Artificial Intelligence,[151] there are numerous other regulators and public undertakings that do not have specific functions relating to data use-cases.

11.   Accordingly we considered that to give a general overview of the extent to which there is legal risk of judicial review challenge, we should first identify the extent to which the UK's current corpus of legislation makes useful specific reference to "Artificial Intelligence" and any of the Concepts.

12.   We found that there are currently very few instances where legislation specifically makes reference to these new technologies, and secondly, that these are only in relation to highly specific contexts.

---

148   We have not systematically searched all Parliamentary Bills; the only Bill of which we are aware which references Artificial Intelligence is the "5 Minute Bill" Artificial Intelligence (Regulation and Workers' Rights) Bill introduced on the 18 May 2023. This is unlikely to proceed.

149   There is a considerable amount of European Union legislation by way of , for instance Commission Decisions and Regulations which refer to Artificial Intelligence. These have not been referenced as the focus of this Snapshot report has been on legislation in the United Kingdom.

150   These are the Competitions and Markets Authority (CMA), Ofcom, the Information Commissioner's Office and the Financial Conduct Authority.

151   However even for them, the extent to which these powers are fully apt for the Concepts is not yet determined.

Public undertakings

13. The second part of our research was to examine whether, and if so to what extent, key bodies operating in the public sphere in the UK, such as government departments, Parliamentary Committees, organisations which carry out government functions, and regulators, were considering the significance of each of these Concepts in their application to society as the 'next generation' Artificial Intelligence.

14. A desk top review of a range of organisations in the public sphere, pre-agreed with the Institute as being particularly relevant, was carried out by Ameer Ismail under our supervision. As relevant material was potentially identified, a deeper examination of its reach was conducted.

15. To maximise the identification of relevant work by these organisations, the research used agreed pre-determined search terms, as set out in **Box B**.

Box B – The pre-determined search terms

"General Purpose Artificial Intelligence"	"Large Language Model"	"Artificial General Intelligence"
"GPAI"	"Large Language Models"	"Frontier Model"
"General Purpose AI"	"LLMs"	"Frontier Models"
"Generative AI"	"LLMS"	"Chat GPT"
"Generative Artificial Intelligence	"Foundation Model"	"ChatGPT"[152]
	"Foundation Models"	"Artificial Intelligence"
	"Foundational Models"	
	"Foundational Model"	

---

152 Note that "Chat GPT" and "ChatGPT" were chosen as search terms due to prevalence of this tool within the mainstream media at present.

16.   The websites of the identified organisations[153] were first interrogated and then, where this yielded poor results, a further search conducted using Google.[154] This method using Google was deployed primarily for the Concepts other than "Artificial Intelligence".[155] The results are summarised in Appendix B.

Summary of findings

17.   Although there are various primary and secondary sources of legislation that use and define "Artificial Intelligence" in fields such as export control, the research (as set out in Appendix A) showed two particularly striking points –

- *No UK based primary or secondary legislation explicitly references the first or 'next generation' Artificial Intelligence in any sphere which relates to how governments or bodies operate in the public sector or more generally make decisions concerning people within the UK.*

- *There is currently no primary or second legislation that explicitly references the 'next generation' of Artificial Intelligence as described in the Concepts.*

18.   The second stage or the research revealed that a number of organisations have been considering the impact of "Artificial Intelligence" since at least 2016 (Appendix B). Some have also been considering the Concepts as noted in **Box C**.

---

153   It was noted during the research that the government website search functions on www.gov.uk both accurately identified the exact search term and could also be filtered by the relevant department. Thus, it was found that searches on the gov.uk domains produced accurate results.

154   Most website search functions were not helpful with more sophisticated terms such as "general purpose AI". Accordingly, Google was interrogated using the following methodology: site:[website domain] ["Concept"] and it was found that this generated fairly accurate results and frequently located information which could not be found by directly searching the website of the relevant body. However, it was a cumbersome approach and we cannot be completely certain that all documents were obtained.

155   Using Google to search websites was found to be less effective when using the term "artificial intelligence" because it usually led to an unmanageable number of hits, often exceeding a few hundred per website domain.

Box C – Public Bodies having some consideration of the Concepts

The Cabinet Office;

Department for Business, Energy and Industrial Strategy;

Department of Education (DoE);

Intellectual Property Office;

Government Office for Science;

Department for Science and Innovation;

The House of Lords Select Committee on Artificial Intelligence;

The House of Lords via its library;

The Communications and Digital Committee of the House of Lords,

Centre for Data Ethics and Innovation (CDEI);

Office for AI;

NHS (England only);

Medicines and Healthcare products Regulatory Agency (MHRA);

Equality and Humans Rights Commission (EHRC);

Information Commissioner's Office (ICO);

OFCOM;

Competition and Markets Authority (CMA);

Financial Conduct Authority (FCA);

Bank of England;

Prudential Regulation Authority (PRA); and

Financial Reporting Council.

19. On the other hand we found that nearly 40 major public bodies do not appear to have made public statements of the implications of the Concepts for their domains at all.[156] These undertakings are set out in Box D.

---

156 Some of these may have made public statements about their engagement generally with Artificial Intelligence, but the research which we have carried out suggests that they have not yet considered the next generation technologies.

Box D – Public Bodies having no apparent consideration of the Concepts

The Ministry of Defence;

The Central Digital and Data Office;

The Surveillance and Biometrics Commissioner;

The Investigatory Powers Commissioner;

The Committee on Standards in Public Life;

The House of Commons Science and Technology Committee;

AI in Weapon Systems Committee;

UK Visas and Immigration in 2013 (taking over the role of the UK Borders Agency);

HM Revenue and Customs;

National Audit Office;

Crown Commercial Service (CCS) (taking over the role of the Government Procurement Service (GPS));

NHS Digital;

Health and Safety Executive;

His Majesty's Inspectorate of Constabulary and Fire & Rescue Services;

The College of Policing;

CAA (Civil Aviation Authority);

NHS Improvement (NHSI) (which became part of NHS England in July 2022);

Director of Public Prosecutions (DPP);

UK Atomic Agency;

Ofqual;

Ofsted;

Care Quality Commission (CQC);

UK Health Security Agency (UKHSA);

OFWAT;

OFGEM;

Office for Nuclear Regulation

General Medical Council (GMC);

Pensions Regulator;

Regulator of Social Housing;

Payment Systems Regulator (PSR);

ORR (Office of Rail and Road);

Office for Professional Body Anti-Money Laundering Supervision (OPBAS);

The Care Inspectorate;

Care Council for Wales (CCW);

Northern Ireland Social Care Council (NISCC);

Scottish Social Care Council (SSCC);

Human Fertilisation and Embryology Authority

Human Tissue Authority (HTA);

Nursing and Midwifery Council (NMC);

General Pharmaceutical Council;

General Medical Council;

Pharmaceutical Society of Northern Ireland (PSNI);

The Advisory Conciliation and Arbitration Service (ACAS).

Conclusions

- Our conclusions are as follows –

  - Having worked for nearly 5 years in this field, we are aware that most of the organisations identified in Box C commenced their thinking about the implications of Artificial Intelligence some time ago. This may therefore explain why they are in the forefront in considering the significance of 'next generation' Artificial Intelligence. Of course, the mere fact that these bodies have been found to have commenced consideration of the significance of 'next generation' Artificial Intelligence, does not imply that such thinking is either complete, advanced or for even apt.

  - The picture revealed by Box D is that there are many major public undertakings in which it appears that there is currently no, or only a very limited, discussion concerning the Concepts is alarming.

  - While it may be debated whether the Concepts are immediately relevant for some of these undertakings identified in Box D, there can be no doubt that, in the very near future, they will increasingly become so, since from our work in this area of technology we well know

that development leads swiftly to deployment.

- We therefore recommend that, at the soonest opportunity, those undertakings which have not as yet thought about these Concepts, should undertake their own landscape review of how they may affect the work they undertake in their particular domain of operation.

- For those undertakings, which have already started this consideration, there is plainly a role to work with those that have not yet done so, to support and encourage this endeavour.

- There is a pressing need for a regular review of the work of these organisations to see how they are keeping abreast of the implications of these new technologies identified in the Concepts, and any future developments.[157]

- We have considered how regularly such a review should be carried out. In our view, for the foreseeable future, the potential impacts of the new technologies on human rights, equality, data privacy and other rights of individuals, is so great that this review should be carried out not less than twice a year, and that a public report of the outcome of that review should be published soon after it is completed.

- To avoid or reduce the risks for regulators and other similar undertakings of a judicial review, in which they are accused of acting ultra vires (outside their powers), some legislative changes will be necessary.

- This could be achieved either directly to the definitions of their statutory functions or through an amendment to the Legislative and Regulatory Reform Act 2006.

Robin Allen KC & Dee Masters

AI Law Consultancy

24 July 2023

---

157   This is not only our view. The DRCF has noted both the need for a clear regulatory approach to Generative AI and its future implications: see "Maximising the benefits of Generative AI for the digital economy" a DRCF Blog, 19 July 2023.

Appendix A: UK Primary and Secondary Legislation

UK Primary and Secondary Legislation referencing "artificial intelligence"

Section 23A(4) of the Enterprise Act 2002 (now repealed) defined "Artificial Intelligence" as "...technology enabling the programming or training of a device or software to use or process external data (independent of any further input or programming) to carry out or undertake (with a view to achieving complex, specific tasks) – (a) automated data analysis or automated decision making; or (b) analogous processing and use of data or information"

Supply and Appropriation (Anticipation and Adjustments) Act 2023, Schedule 1, para 1 makes reference to the Office of Qualifications and Examinations Regulations having a Departmental Expenditure Limit which includes reference to "Exploring, investigating, acknowledging opportunities for innovation, including the use of artificial intelligence to improve the quality of marking in high-stakes qualification".

However no definition of "artificial intelligence" is provided.

Similar provisions as above in the Supply and Appropriation (Main Estimates) Act 2020, 2021, 2022, the Supply and Appropriation (Anticipation and Adjustments) Act 2021, 2022

National Security and Investment Act 2021 (Notifiable Acquisition) (Specification of Qualifying Entities Regulations 2021/1264, Schedule 1, para 3 defines "metamaterials" to include tools that enable artificial intelligence. Equally "other materials" is defined to include "creative artificial intelligence algorithms for material discovery and optimisation".

The purpose of this legislation is to provide the Government with updated powers to scrutinise and intervene in investment to protect national security, as well as to provide businesses and investors with the certainty and transparency they need to do business in the UK (see Explanatory Note).

National Security and Investment Act 2021 (Notifiable Acquisition) (Specification of Qualifying Entities Regulations 2021/1264, Schedule 2, para 1 contains a definition of "core components" which includes "hardware or software enabling sophisticated computational capabilities, including the use of artificial intelligence to process data and data sets received from the sensors and adapt the behaviour of the advanced robotics"

National Security and Investment Act 2021 (Notifiable Acquisition) (Specification of Qualifying Entities Regulations 2021/1264, Schedule 3, para 1.

"artificial intelligence" is defined to mean "technology enabling the programming or training of a device or software to— (i)perceive environments through the use of data; (ii)interpret data using automated processing designed to approximate cognitive abilities; and (iii)make recommendations, predictions or decisions; with a view to achieving a specific objective".

Export of Goods (Control) Order 1985, Schedule 1 contains a definition of artificial intelligence software which includes "software normally classified as expert systems enabling a digital computer to perform functions normally associated with human perception and reasoning or learning".

Various Expert of Goods Orders define software and technology to include artificial intelligence i.e. Schedule 1 of 1985 Order; 1987 Order; 1989 Order.

There are also orders from 1987 and 1989. These orders relate to prohibitions and restrictions on exportation.

Dual-Use and Related Goods (Export Control Regulations) 1995/271 (now repealed) also reference to artificial intelligence software (Schedule 4, para 1).

Appendix B: UK Public Sector

Note: Text **in bold** in this Appendix indicates the source refers to one or more of the new generation concepts. If not in bold then the source refers only to basic concepts of Artificial Intelligence.

Organisation (website accessed, date, Concepts identified)	If resource identified addressing the Concept:
Cabinet Office Website 7.7.23 *Document identified as part of the first stage of the project which was examining legal resources (30.6.23)* • Generative AI	• **Cabinet Office Guidance Note – Guidance to civil servants on use of generative AI (identified through legal research "Generative AI")** • **The use of generative AI in government ("Generative AI").**
Ministry of Defence Website 10.7.23 • Artificial Intelligence	• Defence Artificial Intelligence Strategy, June 2022
Department for Business, Energy and Industrial Strategy Website 7.7.23 Documents also identified as part of the first stage of the project which was examining legal resources (30.6.23) • Artificial Intelligence • Generative Artificial Intelligence • Large Language Models • General Purpose AI	• **Policy paper: Establishing a pro innovation approach to regulating AI", July 2022 (identified through legal research "Generative Artificial Intelligence", "Large Language Models")** • **The Potential Impact of Artificial Intelligence on UK Employment and the Demand for Skills ("General Purpose AI")** • AI Sector Deal

Department for Education

Website

27.6.23

5.7.23

- Artificial intelligence
- Generative AI
- Generative artificial intelligence
- Large Language Models
- Foundation Model
- LLMs
- ChatGPT

Also some documents identified as part of the first stage of the project which was examining legal resources (30.6.23)

- **Departmental Statement - generative artificial intelligence in education (Mar 2023) ("generative AI"; "generative artificial intelligence"; "large language models"; "LLMs"; "foundation model"; "ChatGPT")**
- **DoE Consultative Document: Generative artificial intelligence in education: Call for evidence (identified through legal research against the terms "Generative Artificial Intelligence" and "Large Language Models"**
- Education Secretary addresses BETT 2023 (policy related speech)
- **London tech week speech (Jun 2023) ("generative AI"; "generative artificial intelligence"; "large language models": "ChatGPT"**
- Education Secretary establishes government's forward thinking AI approach (Mar 2023) (policy related speech)
- **Open consultation (Jun 2023) ("generative AI"; "generative artificial intelligence"; "ChatGPT")**
- Review of the online learning and artificial intelligence education market
- Multimillion government investment to fund new AI conversion degrees (Jun 2019)
- 2500 new places on AI and data science conversion courses now open (Jun 2020)
- Prime Minister announces £20 million Institute of Coding (Jan 2018)
- Minister Chris Skidmore will announce plans to fund trials of ground-breaking assistive technology for pupils with special educational needs and disabilities in up to 100 schools and colleges (Jan 2020)
- Early years apps approved to help families kick start learning at home (Feb 2020)
- International Education Strategy: global potential, global growth (Feb 2021)
- Thousands of working adults will be offered free courses to rapidly upskill in STEM sectors including engineering, digital and construction (Oct 2021)

Centre for Data Ethics and Innovation (CDEI)

Website

3.7.23

Revisited 5.7.23

- Artificial intelligence
- Generative AI
- LLMs / LLMS

"Generative artificial intelligence" led to one hit but reference not found.

- Terms of reference 2021- 23 – Advisory board of the CDEI.
- CDEI report on 'Industry Temperature Check' (Dec 2022)
- CDEI portfolio of AI assurance techniques (June 2023)
- CDEI research on AI governance (March 2023)
- Find out about AI assurance techniques (case studies)

E.g.

- Case study from Fairly AI ("LLMs")
- Case study from Logically AI on misinformation
- Case study from Trilateral Research on AI used to safeguard children
- Case study from the Food Standards Agency
- Case study from the British Standards Institution
- Case study from The Alan Turing Institute
- Case study from Qualitest on NHS England
- Case study from Mind Foundry on fraud detection in insurance
- Case study from Best Practice AI – for an AI-enabled medical symptom checker – Jun 2023
- Case study from Nvidia on credit risk management
- Case study from BABL AI on conducting third-party audits for automated employment decision tools
- Case study from Shell
- **Case study from Citadel ("generative AI"; "LLMs")**
- Advisory Board of the Centre for Data Ethics and Innovation (transparency data - April 2022)
- Independent report on Covid-19 repository and public attitudes - March 2021.
- CDEI Advisory Board – register of interests (transparency board) - Nov 2022
- Guidance for organisations using the Algorithmic Transparency Recording Standard – Jan 2023.
- Algorithmic Transparency Recording Standard Hub (Jan 2023)
- CDEI Privacy Notice re. Collecting use case submissions of AI systems
- CDEI – Two year review – Jul 2021
- Two year strategy (Mar 2019)
- CDEI final report of review into bias in algorithmic decision-making (Nov 2020)
- Independent Report on CDEI's approach to the governance of data-driven technology
- CDEI report on public attitudes to Data and AI (Wave 1 – Dec 2021)
- CDEI report on public attitude to data and AI (Wave 2 – Nov 2022)
- CDEI report on public attitudes to digital regulation
- CDEI roadmap to building an AI assurance ecosystem in the UK (Dec 2021)
- Transparency data – research into algorithmically driven music recommendation systems
- Report by 'BritainThinks' on algorithmic transparency in the public sector (Jun 2021)
- Report by BritainThinks on public trust in data use (Dec 2021)
- Interim report into online targeting and bias in algorithmic decision-making (Jul 2019).
- CDEI Bias Review – Summary of responses (Oct 2019)
- CDEI final report and recommendation on social media targeting (Feb 2020)
- CDEI Report on facial recognition technology (May 2020)
- CDEI report on responsible access to demographic data to make AI systems fairer (Jun 2023)
- Report on Self-driving vehicles (Aug 2022).
- CDEI's second edition of its AI Barometer (Press release - Dec 2021)
- CDEI AI Barometer 2020; and 2021.
- CDEI Snapshot Paper on Deepfakes and Audiovisual Disinformation - Sep 2019.
- CDEI report on role of AI in addressing misinformation on social media (Aug 2021)
- **UK launches data reform to boost innovation, economic growth and protect the public (Sep 2021) ("gpai" - although refers to GPAI the organisation)**
- Six lessons for AI Assurance profession to leave from other domains – part one: how can certification support trustworthy AI?, 12 July 2023. There are six parts to this series in total.

Office for AI

Website

3.7.23

Revisited 5.7.23 and 10.7.23

- Artificial intelligence
- GPAI
- General purpose AI
- Generative AI
- Generative Artificial Intelligence
- Large language Model/s
- LLM
- Foundation models
- Artificial general intelligence
- ChatGPT

- AI sector study 2022
- **Achievements under the AI Sector Deal (May 2021) ("GPAI")**
- A guide to using artificial intelligence in the public sector
- Policy paper on AI Sector Deal (to boost UK's global position in developing AI tech).
- Turing AI Fellowships Guidance
- Guidance on Understanding AI in the public sector
- Guidelines for AI procurement
- Secretary of State letter to DCRF on a pro innovation approach to regulating AI (Mar 2023)
- Analysis of hypothetical AI regulatory options for the UK (Mar 2023)
- Guidance on how to manage a project which uses artificial intelligence
- Guidance on using AI in the public sector (last updated Oct 2019)
- Case studies referred to therein e.g.
- Signalling company to help trains
- DFID estimating populations using satellite information
- Guidance on using AI in the public sector (last updated Jan 2020)
- Guidance on understanding AI ethics and safety (Jun 2019)
- Declaration of USA and UK on cooperation in AI research and development
- **Global Partnership on AI (GPAI) statement (Jun 2020) ("GPAI")**
- Guidance on planning and preparing for implementing AI (Jun 2019)
- Guidance to help assess if AI is the right technology for your challenge (Jun 2019)
- Research on Understanding UK AI Research & Development and the Role of Standards (May 2022)
- Research on AI activity in UK businesses (Jan 2022)
- **White Paper on pro-innovation approach to AI regulation (Jun 2023) ("GPAI"; "general purpose AI"; "generative AI"; "generative artificial intelligence"; "large language models"; "LLMs"; "Foundation models"; "ChatGPT")**
- **See Consultation description (Mar 2023)**
- **See Policy Paper on an overview of UK's emerging approach dated Jul 2022 ("GPAI"; "Large Language Model(s)"; "LLM")**
- Report on AI Roadmap - to help govt. strategic direction on AI – Jan 2021
- National AI Strategy promotional material (Sept 2021)
- Research on AI-assisted patent prior art searching (Apr 2020)
- Guidance on ethical use of AI decision-making systems (May 2021)
- Ipsos UK artificial intelligence sectoral analysis survey (Oct 2022)
- Research on the UK AI labour market (May 2021)
- **National AI Strategy (Dec 2022) ("artificial general intelligence")**
- **Guidance ("GPAI" - although referring to the organisation; "general purpose AI"; "artificial general intelligence").**
- See also: AI Action Plan (Jul 2022)
- Evaluation of the Industry-funded Masters in AI programme
- EY analysis to assess extent of data foundations and AI adoption (Aug 2021)
- **New strategy to unleash the transformation power of Artificial Intelligence (Mar 2021) ("gpai")**

**Central Digital
and Data Office**

Website

3.7.23

- Artificial Intelligence

- Number of joint publications already cited above w/ OFAI e.g.:
- A guide to using AI in the public sector (joint w/ OFAI) (Jun 2019)
- Guidance on how to manage a project which uses artificial intelligence (joint w/ OFAI) (Oct 2019)
- Guidance on using AI in the public sector (last updated Oct 2019) (Joint w/ OFAI)

[not all joint publications already cited above are repeated]

- Guidance on integrate and adapt technology (updated Mar 2021)
- Guidance for public sector orgs on how to use data appropriately (last upd: Sep 2020)
- Roadmap – transforming for a digital future (Jun 2022)
- Policy paper on technology innovation in government survey (Aug 2018)
- Government Technology Innovation Strategy (Jun 2019)
- Government Functional Standard on Digital Data and Technology in govt (Mar 2023)
- Guidance on digital, data and technology essentials for senior civil servants (May 2022)
- Government technology standards and guidance – Aug 2016
- AI tool to assist with Food Hygiene Rating Scheme (Jul 2022)

**Department for Science,
Innovation and Technology
(DSIT)**

Website

10.7.23

- Artificial intelligence
- Large Language Model

- AI Council
- Newsletter (July 2023)
- Artificial Intelligence Sector Study, 2022
- **The UK Science and Technology Framework ("Large Language Model")**

**Government Office
for Science**

Website

10.7.23

- Artificial intelligence
- Generative AI
- Foundational Model
- Foundation Model

- **Rapid Technology Assessment: Artificial Intelligence, 2023 ("Generative AI", "foundational model", "foundation model").**

Surveillance and Biometrics Commissioner	• Surveillance camera code of practice (2021)
	• Plus Amended version (2022)
[Website](#)	• Policy paper – DCMS consultation: "Data: a new direction": response by the Commissioner"
30.6.23	• Letter from the Biometrics and Surveillance Camera Commissioner to Lucy Allan MP about cameras around MPs' home addresses
Revisited 5.7.23	• News story on Drones, DNA losses and mission creep feature in OBSCC annual report
	• Annual Reports dated 2021 – March 2022.
• Artificial Intelligence	• Transparency date – Letter to Biometrics and Surveillance Camera Commissioner from Hikvision 22 June 2021
• AI	• Letter from BSCC to Dahua 11 April 2022
	• The Biometrics and Surveillance Camera Commissioner's submission to the DPDI Bill Committee
	• Biometrics and Surveillance Camera Commissioner speech at NPCC CCTV Conference March 2022
	• Surveillance Camera Commissioner Newsletter – May 2021

Investigatory Powers Commissioner	• Technological Advisory Panel
	• Fives Eyes intelligence oversight meeting (Oct 2019)
[Website](#)	• IPCO attendance at Essex University's conference in May 2018.
3.7.23	• Annual Report 2021
Revisited 5.7.23	• Annual Report 2020
• Artificial Intelligence	

The Office of the Police and Crime Commissioner	• Not manageable since there is no single website for the OPCC.

Intellectual Property Office	• Intellectual property and investment in Artificial Intelligence
[Website](#)	• Government response to call for views on artificial intelligence and intellectual property
10.7.23	• **Guidance: The Government's Code of Practice on copyright and AI ("Generative AI")**
• Artificial intelligence	
• Generative AI	

The Committee on Standards in Public Life (CSPL) Website 3.7.23 Revisited 5.7.23 • Artificial Intelligence (15 hits)	• AI and Public Standards – Terms of Reference • Artificial Intelligence and Public Standards (Mar 2019) • Announcement of review into AI and Public Standards (Mar 2019) • Report on Artificial Intelligence and public standards (Feb 2020) • Public polling on attitudes towards AI (Feb 2020) • CSPL submission to Science and Technology Committee inquiry on Governance of AI • Written evidence submitted to the Committee's review into artificial intelligence and public standards (Feb 2020) • CSPL Submission to Cabinet Office Consultation on Public Procurement (Submission echoes the conclusions of its Feb 2020 report on AI). • Focus group on AI and public standards • Govt response to the Committee on Standards in Public Life's 2020 report on AI and public standards, May 2021 • Artificial Intelligence and Public standards: regulators survey and responses, Jan 2021 • Annual Report 2019 - 20; and 2018 - 19
The House of Commons Science and Technology Committee Website Inquiries Reports News (no search bar so used Google, see here) 3.7.23 Revisited 5.7.23 • Artificial Intelligence	• Governance of artificial intelligence • Robotics and artificial intelligence inquiry • Robotics and artificial intelligence – Fifth Report of Session 2016 - 17 • See Response News: • MPs to examine regulating AI in new inquiry (2022) • Alan Turing Institute writes to Chair of Committee (2016) • MPs call for research sector reforms to address concerns with reproducibility of science (2023) • Science and Technology Committee is launching a new inquiry into the use of algorithms (2017) • Quantum technologies inquiry launched (2018)
The House of Lords Select Committee on Artificial Intelligence Website 3.7.23 • Artificial Intelligence • Artificial General intelligence	• **Artificial Intelligence Committee AI in the UK: ready, willing and able? Apr 2017 ("artificial general intelligence")** • Govt. Response to Report (Jun 2018) • Plus Cover Letter from Sam Gymah MP and Margot James MP – June 2018 News: • UK can lead the way on ethical AI • What should the government do about artificial intelligence (Dec 2017)
The House of Lords Select Committee – The Communications and Digital Committee Website 10.7.23 • Generative AI • Large Language Models	• **Launch of an inquiry into large language models ("Generative AI", "Large Language Models").**

House of Lords Library Website 18.7.23 • Artificial Intelligence • Generative AI • Chat GPT	• **Artificial intelligence: Development, risks and regulation ("Generative AI", "Chat GPT) (18 July 2023)**
AI in Weapon Systems Committee Website 3.7.23 • Artificial intelligence	• The Inquiry • How should autonomous weapons be developed, used and regulated? (Mar 2023)
The All-Party Parliamentary Group on Artificial Intelligence (AIPPG) Website	N/A [website too difficult to navigate]
UK Border Agency Website 3.7.23 Revisited 5.7.23	No concepts identified
UKBA was replaced by UK Visas and Immigration in 2013 Website 3.7.23 Revisited 5.7.23 • Artificial Intelligence	• Work in the UK as a leader in digital technology (Global Talent visa) • Guidance on sponsoring a worker • Guidance on Technical or business skills covered by Tech Nation • List of authorised endorsing bodies that were able to issue endorsements for start-up visas before 14 April 2023 • Press release – unlimited visa offer to world's best and brightest (Jan 2020)
HM Revenue and Customs Website 3.7.23 Revisited 5.7.23 • Artificial Intelligence	• HMRC Internal Manual – Corporate Intangibles Research and Development Manual • Case Studies • HMRC Corporate report – annual report and accounts 2021 to 2022 • The future of customs declarations – call for evidence • Policy paper –No Safe Havens 2019: assisting compliance & No Safe Havens 2019: responding appropriately • HMRC appoints Chief Digital and Information Officer (Sep 2020) • Tax Administration Framework Review – information and data (Apr 2023) • RCDTS Annual Report and Accounts 2021 - 2022

National Audit Office

Website

3.7.23

Revisited 5.7.23

- Artificial Intelligence

- How to audit artificial intelligence models (Jun 2021)
- Financial services regulation – adapting to change (Work in progress)
- UK Research and Innovation's management of the Industrial Strategy Challenge Fund - Report
- Reforming the civilian workforce
- Insight – Right data, right place, right time

Government Procurement Service (GPS)

Closed. Now part of the Crown Commercial Service.

Crown Commercial Service (CCS)

Agreements

Suppliers

Products and services

News

[NB search feature in website did not search defined terms. Used this method instead.]

4.7.23

- Artificial intelligence

Agreements:

- Agreement on Artificial Intelligence
- Agreement on Big Data and Analytics
- Agreement on Spark DPS
- Agreement on Outsourced Contact Centre and Business Services
- Agreement on Traffic management Technology (Expired)
- Agreement on Cyber Security Services 3
- Agreement on Transport Technology & Associated Services
- Agreement on Permanent Recruitment Solutions (Expired)
- Network Services 2
- Automation Marketplace DPS
- Data and Application Solutions
- Public Sector Resourcing

Suppliers

- N/A
- Products and services
- Artificial intelligence
- Automation
- Technology
- Digital Future
- Technology MoU

News

- Artificial intelligence and automation for the public sector (May 2021)
- Discover our new AI dynamic purchasing system (Sep 2020)
- Innovative new automation marketplace DPS goes live (Mar 2020)
- Our new first of its kind agreement specifically for big data and analytics services and software goes live (Sept 2022)
- Breckland Council expect to make 30% efficiency savings by using chatbots (Jul 2021)
- Role of data and analytics in digital transformation (Oct 2022)
- Chatbots and virtual assistants for local authorities (Jun 2021)
- Supporting the public sector to improve data use, drive efficiency and improve services (Sep 2022)
- Spark's new technology innovation offerings (Sep 2019)
- Closing the digital skills gap to support digital transformation (Dec 2022)
- Crown Commercial Service and Google Cloud sign MoU to boost cloud innovation (Jun 2020)
- Maximising efficiency from contact centres and business services (Dec 2021)
- Security of sustainability (Oct 2021)
- CCS and Microsoft Azure agree new pricing arrangement (May 2020)

NHS England, Scotland and Wales and NHS X

- England
- Scotland
- Wales

[NHS X merged with NHS digital and incorporated into NHS England: see here]

4.7.23

England

- Artificial intelligence
- Generative AI
- Large Language Models
- Foundation model

Wales

- None

Scotland

- None

England

- First wave of Test Beds; Wave 2
- Delivering plan for tackling the Covid-19 backlog of elective care
- Artificial in Health and Care Award
- Interim NHS People Plan: the future medical workforce
- Evolving to meet a changing world
- Children and young people's elective recovery toolkit (
- Virtual clinics in Highly Specialised Services (HSS): guidance for services supporting patients with rare and complex and multi-system disorders (Mar 2023)
- Topol Review on Preparing the healthcare workforce to deliver the digital future (Feb 2019)
- Science in healthcare (Mar 2020)
- CVDR and Diabetes Clinical Networks (2021 – 2022 report)
- Delivering a net zero NHS
- Five year review – Harnessing technology and innovation
- **NHS Long term workforce plan (Jun 2023) ("generative AI"; "Large Language Models"; "Foundation model")**

Case studies:

- Artificial intelligence & skin cancer diagnosis (Feb 2023)
- How AI is supporting referrals (Aug 2018)
- Free transport reduces was not brought rates for children at Midlands trust
- WNB AI tool (May 2023)
- Paediatric accelerator (May 2023)
- National Stroke Service Model (May 2021)
- Accelerating genomic medicine in the NHS (Oct 2022)

News:

- NHS pilots AI software to cut missed hospital appointments (Feb 2023)
- NHS uses AI to speed up stroke care (May 2020)
- AI software in Neuroscience for Stroke Decision Making Support
- NHS aims to be a world leader in AI and machine learning within 5 years (Jun 2019)
- NHS expands high-res skin imaging (Jun 2023)
- NHS cuts elective backlog with longest waiters down a quarter in one month (Feb 2023)
- NHS in east cuts number of longest waits by almost quarter in a month (Feb 2023)
- North East and Yorkshire health and care heroes (Jun 2023)
- NHS backs pioneering new cancer innovations set to transform cancer diagnoses in England (Feb 2022)
- Update on outpatient transformation (NW)
- Modern screening can be more personalised and convenient to save lives says new report (Oct 2019)

Blogs:

- Innovation: the golden thread empowering the NHS for a bright future (May 2023)
- 75 years young (Mar 2023)
- Knowing what good looks like (Dec 2018)
- Delivering the digital future (Aug 2018)
- Our role in developing digital capabilities (Apr 2019)

NHS Digital

Digital

Website

4.7.23

- Artificial intelligence

- Solution assurance
- Social Care Programme
- Benefits of data quality (Part of Data quality assurance framework for providers Part 1)
- Benefits of data quality (Part of Data Quality Assurance Framework for Providers Part 2)
- All digital services need to be delivered sustainably (Part of NHS architecture principles)
- Digital Sustainability (Part of Sustainability Annual Report 2019-20)
- Public perceptions of NHS data use: Rapid literature review (Jul 2021)
- Delivering digital services for NHS Test and Trace
- Sustainability Annual Report 2020-21 (Summary of highlights)
- Digital Social Care Pathfinders Programme 2019 - 21
- City of Wolverhampton Council project
- Worcestershire County Council project
- London Borough of Islington
- Friends of the Elderly
- Google Cloud training (esp. this)
- AWS Cloud Skills – partnership with AWS (I.e. Amazon) to provide AWS cloud courses
- AWS NHS migration case studies (esp. This)

Case studies:

- Technomed
- Free text data for adult social care prevention and intervention

Blogs:

- Welcoming new health data research hubs (Sep 2019)
- Data saves lives (Mar 2020)
- Regulating AI in health and care (Feb 2020)
- Celebrating the lasting legacy (Sep 2021)
- Covid-19 digital readers: can you see the line? (Nov 2022)
- 5 key themes for the future of the NHS Spine (Dec 2020)
- Tell us where data-driven technology is heading (May 2019)
- Are the bots taking over? No, but we're going to put them to good work (Jul 2018)
- What next for data in the NHS (May 2019)
- Past present and future of innovation in hospital and secondary care (Jul 2018)
- Past present and future of innovation to support and empower the patient (Jul 2018)
- Past present and future of healthcare data innovations (2018)

News:

- Partnership with Great Ormond Street Hospital for new technology innovation centre (Oct 2018)
- DRIVE partnership announcement (Sep 2021)
- Digital Social Care Pathfinder helps charity integrate acoustic monitors (Nov 2020)
- NHS Digital welcomes funding for joint project with MHRA on creating synthetic devices (Oct 2018)
- High-tech 'Find and Treat' service (Mar 2022)
- Digital revolution in social care (Aug 2021)
- Digital tools to help improve access to healthcare for patients with hearing and sight loss (Aug 2019)
- NHS Digital comment on Reform report into AI in healthcare (Jan 2018)
- How Moorfields is using AutoML to enable clinicians to develop machine learning solutions (Sep 2019)

Health and Safety Executive Website 4.7.23 • Artificial Intelligence	• Industrial use of safety related expert systems (2000) • Foresight report 2016 • Foresight Annual Report 2017/18 • Evaluation into the success of occupational health and safety regulators and organisations use of expert systems (2006)
His Majesty's Inspectorate of Constabulary and Fire & Rescue Services Website 5.7.23 • Artificial intelligence	• HMICFRS Strategy 2021 - 25 • State of policing: The Annual Assessment of Policing in England and Wales 2021 (Mar 2022) • See annual assessment 2020 • Annual assessment 2019 • Annual assessment 2018 • An inspection of how well the police and National Crime Agency tackle the online sexual abuse and exploitation of children (Apr 2023) • Force Management statement: Template for forces (Oct 2022) • Force management statements 3: guidance and template (Mar 2023) • Observations on second Force Management Statement (Mar 2023) • An inspection into how well the police and other agencies use digital forensics in their investigations (Dec 2022) News: • HMCIC criticises lack of upstream public investment (Sep 2019) • Profound and far reaching policing reform urgently needed (Jul 2019) • Planning failures in policing risk public safety (Jun 2018) • Speech by Sir Thomas Winsor on 10 Sep 2019 ('in policing there is virtually no use of AI') • See speech on 24 June 2021 • [NB PEEL reports into individual Police departments ignored]
College of Policing Website 5.7.23	• No concepts identified
CAA (Civil Aviation Authority) Website 5.7.23 • Artificial intelligence	• The innovation team • The CAA Cyber Programme • World ATM Congress (Mar 2019) • Annual Report 2019/20
NHS Improvement (NHSI) NHSI became part of NHS England in July 2022. Website 5.7.23 Artificial intelligence	• Healthcare UK annual review 2019 to 2020 (Oct 2020)

Director of Public Prosecutions (DPP)

NB unable to find specific website so relied on https://www.cps.gov.uk/about-cps

5.7.23

- Artificial intelligence

News

- DPP Max Hill's speech to the Cambridge Symposium on Economic Crime (Sep 2022) (Speech related to policy)
- DPP Max Hill Speech at King's College London (Mar 2022) (Speech related to policy)
- DPP Max Hill speech to Heads of Prosecuting Agencies Conference (Jul 2022)

UK Atomic Agency

10.7.23

- Artificial Intelligence

- Industrial Metaverse essential for UK fusion energy development, June 2023

The Information Commissioner's Office (ICO)

Website

3.7.23

Revisited 5.7.23

DRCF – 19.7.23

- Artificial intelligence (172 hits – only the clearly relevant Resources were recorded)
- Generative AI
- Generative Artificial Intelligence
- GPAI
- Large Language Models
- LLMs
- ChatGPT

- Guidance on AI and data protection
- Guidance on Explaining decisions made with AI
- Guide to accountability and governance (GDPR)
- Guidance on research-related processing (GDPR)
- AI and data protection risk toolkit
- Data analytics toolkit
- Regulatory Sandbox (NB – number of previous case studies from Reg Sandbox involving AI – not recorded)
- Our work on artificial intelligence
- Information Rights Strategic Plan 2017 - 2021
- **ICO's response to Government's AI White Paper (2023) ("generative AI"; "GPAI")**
- **ICO's response to HOC S&T Committee inquiry into the governance of AI ("gpai")**
- ICO summary of response to the consultation on ICO guidance on the AI auditing framework.
- ICO call for views on employment practices and data protection
- How to use AI and personal data appropriately and lawfully
- ICO's response to DCMS consultation "Data: a new direction" 2021
- Digital Regulation Cooperation Form – first annual plan of work (Mar 2021) (news)
- Project ExplAIn interim report 2019
- Big Data, artificial intelligence, machine learning and data protection Report (2017)
- MeVitae AI Data Protection Audit Report – Nov 2021
- Guidance on Right to be informed (GDPR)
- What common issues might come up in practice? (re. Right to be informed Guide (GDPR)).
- Rights related to automated decision making including profiling (GDPR)
- Guidance on Automated decision-making and profiling (GDPR)
- A guide to individual rights (GDPR)
- Guidance on Video surveillance (GDPR)
- FAQs on the 15 standards of the Children's code
- Cf. Age assurance – estimating or verifying the age of service users
- Guide to ICO Audit: AI Audits
- Guide on examples of processing 'likely to result in high risk'
- **Previously asked questions: on innovation advice ("generative ai")**
- Roundtable of G7 Data Protection and Privacy Authorities 7 – 8 Sep 2021
- Summary of responses to the consultation on ICO/Turing draft guidance on Explaining AI decisions
- Data Protection Bill, HoC Second Reading – IC's briefing
- Data Protection Bill, House of Commons Committee – IC's written evidence
- Annex - Data Protection Bill, HoL Committee Stage
- Annex II Data Protection Bill, HoL Report Stage – IC's briefing
- The Information Commissioner's position paper on the UK Government's proposal for a trusted digital identity system Apr 2021
- ICO innovation hub project report
- ICO Consultation on Age Appropriate Design Code 2019
- ICO's Opinion: Age Assurance for the Children's Code
- Paper on Democracy Disrupted? - Jul 2018
- The Future of Political Campaigning Jul 2018
- Employment practices and data protection: information about workers' health – Oct 2022
- ICO's Annual Report and Financial Statements 2017-18
- 2018 - 19
- 2019 - 20
- Annual Report and Financial Statements 2020 - 21
- 2021 - 22
- Measurement of Age Assurance Technologies (from Age Check Certification Services) 2022 and 2021 (accs 2)

Thought pieces:

- Opinion on Age Assurance for the Children's Code 2021
- Opinion on use of live facial recognition technology by law enforcement in public places 2019
- Opinion on use of live facial recognition technology by law enforcement in public places 2021

Blogs

- Guest blog: Working for a regulator like the ICO
- **Generative AI: eight questions that developers and users need to ask ("generative ai"; "generative artificial intelligence"; "large language models"; "LLMs"; "ChatGPT")**
- Reflecting on the first year of the 'Explaining decisions made with AI' guidance
- Making our employment guidance work for you
- Why addressing AI-driven discrimination is so important
- What's next for data ethics?
- Call for views: data protection and employment practices

News:

- DPPC 2022 news bulletin
- **Don't be blind to AI risks in rush to see opportunity – ICO reviewing key businesses' use of generative AI ("generative AI")**
- **John Edwards, IC, keynote speech May 2023 ("generative AI")**
- **ICO urges organisations to harness the power of data safely by using privacy enhancing technologies ("generative ai")**
- "Maximising the benefits of Generative AI for the digital economy" (19 July 2023, DRCF)

Equality and Human Rights Commission Website 3.7.23 Revisited 5.7.23 - Artificial intelligence - General purpose AI - Large Language Models - LLMs - Foundation models - ChatGPT	- EHRC has made tackling discrimination in AI a major strand of its new three-year strategy (Sep 2022) - Strategic plan 2022 – 2025 (Mar 2022) - Artificial intelligence in public services guidance - Guidance on how PSED applies when a public body uses AI (Sep 2022) - AI safeguards 'inadequate' (news – Jun 2023) - AI checklist for public bodies in England (Sep 2022) - Business plan 2023- 2024; and 2022 - 2023 - Litigation enforcement policy - Types of cases we are interested in hearing about - **EHRC's response to AI regulation White Paper (Jun 2023) ("general purpose AI"; "Large Language Models"; "LLMs"; "Foundation Models"; "ChatGPT")** Blog posts: - Monitoring human rights in a digital world: sharing our work internationally (Dec 2022) - Black History Month: how EHRC acts against race discrimination (Oct 2022) - UK's civil and political rights record under review: challenges of new digital technologies (Mar 2020) - Inclusive Britain: Responding to the Government's progress updates (May 2023)
Ofqual Website 4.7.23 - Artificial intelligence	- Annual Report and Accounts 2020 to 2021 - Research - A review of approaches to assessing writing at the end of primary education (Mar 2019) - Research - Delivery and award of vocational and technical qualifications in 2022 (Dec 2022) - Regulation – Vocational and technical qualifications Covid 19 Guidance (Aug 2021)

Ofsted	• Research review series: computing (May 2022)
Website	• Amanda Spielman speech at Annual Apprenticeship Conference 2020
4.7.23	• Ofsted Annual report 2018/19: HMCI commentary
• Artificial intelligence	

Care Quality Commission (CQC)	• Evaluation of CQC's regulatory sandboxing pilot (May 2022)
Website	• CQC's strategy from 2021: equality impact assessment (Apr 2022)
4.7.23	• CQC Partnerships
	• Using machine learning in diagnostic services (Mar 2020)
• Artificial intelligence	• Ionising Radion Regulations – Notify us (Apr 2023)
	• Guidance – Scope of registration: Regulated activities (May 2022)
	• Buckinghamshire NHS Trust CQC Inspection (Jul 2022)
	• The world of health and social care is changing. So are we. (Apr 2022)
	• Memorandum of understanding – NICE (May 2022)
	News:
	• NICE and CQC release updated joint working agreement (May 2021)
	• New CQC strategy champions regulation driven by people's experiences of care (May 2021)
	• CQC's strategy from 2021 (Nov 2022)

Medicines and Healthcare products Regulatory Agency (MHRA)

Website

4.7.23

7.7.23

- Artificial Intelligence
- Large Language Models

- Guidance on Software and Artificial Intelligence (AI) as a Medical Device (Apr 2023)
- Guidance – the Target Development Profile Toolkit (Feb 2022)
- Software and AI as a Medical Device Change Programme (Jun 2023)
- Medicines and Healthcare Products Regulatory Agency: Corporate Plan 2023 to 2026 (Jul 2023)
- Medicines and Healthcare Products Regulatory Agency annual report and accounts 2021/22 (Jul 2022)
- 2019/20
- Medicines and Healthcare products Regulatory Agency Delivery Plan 2021-2023
- Medicines and Healthcare products Regulatory Agency Business Plan 2020 to 2021 (June 2020)
- Putting patients first: Delivering our priorities - Delivery Plan 2021-2023 updates for year two (Oct 2022)
- Electronic Prescribing and Medicines Administration Systems: report adverse incidents on a Yellow Card (JAN 2023)
- Guidance for manufacturers on reporting adverse incidents involving Software as a Medical Device under the vigilance system (May 2023)
- Guidance on Good Machine Learning Practice for Medical Device Development
- Guidance on Regulatory status of software used in the diagnosis, treatment and management of patients with Covid-19 (Dec 2021)
- Consultation on the future regulation of medical devices in the UK (June 2022)
- Case study - Horizon Scanning Case Study: Guiding principles that can inform the development of Good Machine Learning Practice (Dec 2022)

Blog:

- **Large Language Models and software as a medical device, March 2023 ("large language model")**

Press release:

- Global AI challenge to investigate veterans' cardiovascular health launched (May 2023)
- New synthetic datasets to assist Covid-19 and cardiovascular research (Jul 2020)
- Transforming the regulation of software and artificial intelligence as a medical device (Sep 2021)
- UK to strengthen regulation of medical devices to protect patients (June 2022)
- MHRA puts delivering for patients at the heart of its Delivery Plan 2021-2023 (Jul 2021)
- Projects lay the groundwork for a future of robolawyers and flying cars (Oct 2018)
- MHRA to receive £10m from HM Treasury to fast-track patient access to cutting-edge medical reports (Mar 2023)
- UK, USA and Canadian regulators identify 10 guiding principles to be addressed when medical devices use AI or machine learning software (Oct 2021)

OFCOM

Website

5.7.23

DRCF – 19.7.23

- Artificial intelligence
- Generative AI
- Large Language Models
- ChatGPT

Reports and discussion

- Supporting innovation in the 100-200GHz range (Oct 2020)
- Online Nation 2021 report
- Spectrum management strategy for the 2020s (Jul 2021)
- Ofcom's plan of work (Mar 2021)
  - See Dec 2021 (proposed plan of work - consultation doc)
  - See Mar 2022
  - Spectrum roadmap (Mar 2022)

Research:

- Review of latest developments in the Internet of Things (IoT) (Mar 2017)
- Use of AI in online content moderation (Jul 2019)
- Online market failures and harms (Oct 2019)
- Automated content classification systems (Jan 2023)
- Enabling 5G in the UK (Mar 2018)
- Case study monitoring political advertising (Feb 2021)
- Understanding online hate by The Alan Turing Institute (commissioned and funded by Ofcom) (Feb 2021)
- **Tracking abuse on twitter against football players by The Alan Turing Institute (commissioned by Ofcom) (2022) ("Large Language Models")**
- Supporting the UK's wireless future – Spectrum management strategy 2020
- **Technology Futures – Spotlight on the tech shaping comms for the future (Jan 2021) ("Generative AI")**
- See Jul 2021

News:

- Big tech developments to watch out for in 2022
- Five themes from 2022 Mobile World Congress
- Seven in ten PL footballers face Twitter abuse (Aug 2022)
- UK's digital watchdogs take a closer look at algorithms as plans set out for year ahead (Apr 2022)
- **What generative AI means for the communications sector (Jun 2023) ("Generative AI"; "ChatGPT")**
- **How we're responding to renewed interest in immersive technologies (Jun 2023) ("Generative AI")**
- **Top tech trends to look out for in 2023 (Jan 2023) ("ChatGPT")**

Bulletin:

- Note to broadcasters on synthetic media (including deepfakes) (Apr 2023)

Blog:

- **"Maximising the benefits of Generative AI for the digital economy" (19 July 2023, DRCF)**

Competition and Markets Authority (CMA)

Website

5.7.23

DRCF – 19.7.23

- Artificial intelligence
- GPAI (but with ref to https://gpai.ai/)
- Generative AI
- Generative artificial intelligence
- Large language model
- Large Language models
- Foundation Model (4 hits)
- Foundation Models (same hits as above)

No other Concepts were identified [NB. "foundational model" had hits but the specific reference to the Concept could not be i dentified].

Reports and policy papers:

- CMA Annual plan 2019/20 (Feb 2019)
- CMA's Digital Markets Strategy Feb 2021 refresh
- DRCPF Plan of Work 2021 to 2022 (Mar 2021)
- **The CMA's response to govt's White Paper 'AI regulation: a pro-innovation approach' (Jun 2023) ("Foundation model")**

Research:

- Modernising consumer markets green paper: CMA response to govt. Consultation (Jul 2018)
- AI in the boardroom 2019 (updated Mar 2022)
- Software validation and AI in finance 2021 (BoE) (Mar 2022)
- AI – A worldwide overview of AI patents 2019 (IPO) (Mar 2022)
- **AI Foundation Models: initial review (May 2023) ("Foundation Models")**
- Pricing algorithms research, collusion and personalised pricing (Jun 2018)
- Algorithms: how they can reduce competition and harm consumers (Jan 2021)
- Consultation outcome: algorithms, competition and consumer harm: call for information (Jun 2021)
- Findings from the DRCF Algorithmic Processing workstream – Spring 2022 (Sep 2022)
- **The benefits and harms of algorithms: a shared perspective from the four digital regulators (Sep 2022) ("GPAI" but with reference to gpai.ai)**
- Explaining why the computer says 'no' (FCA) (Aug 2022)
- Comparing minds and machines: implications for financial stability 2021 (BoE) (Mar 2022)
- AI-assisted patent prior art searching feasibility study 2020 (IPO) (Mar 2022)

News:

- UK competition enforcement – where next? (Nov 2017)
- Future of competition enforcement in the UK (Nov 2018)
- Digital Markets: using our existing tools and emerging thoughts on a new regime (Oct 2020)
- **CMA launches review of AI models (May 2023) ("Generative artificial intelligence"; "foundation models"; "large language model"; Large language models")**
- UK's digital watchdogs take a closer look at algorithms (Apr 2022)
- Joined up approach to digital regulation (Mar 2021)
- **Competition and Innovation: a priority for the CMA (Jun 2023) ("Generative AI"; "Foundation model")**
- **"Maximising the benefits of Generative AI for the digital economy" (19 July 2023, DRCF)**

Financial Conduct Authority (FCA)

Website

5.7.23

DRCF – 29.7.23

- Artificial intelligence
- ChatGPT

- Consultation on AI (Oct 2022)
- Business plan 2023/24
- Digital sandbox (Apr 2023)
- Regulatory Sandbox
- 2017 update on Reg Sandbox
- FCA Innovation Hub
- Data analytics and AI (Oct 2020)
- Case Study: Innovation can benefit us all (Feb 2023)
- Business plan 2022/23 (Apr 2022)
- 'Publish a joint discussion paper with the Bank of England on artificial intelligence in financial services'
- **FCA Employee Handbook pg.20 ("ChatGPT")**

Publications

- Insights from the 2020 Cyber Coordination Groups (Apr 2021)
- FCA Research agenda (April 2019)
- AI Public-Private Forum: Final Report (BoE and FCA) (Feb 2022)
- Research Note: Machine learning in UK financial services 2022
- Artificial Intelligence and Machine Learning (Oct 2022) (BoE and FCA)

News and speeches (examples):

- Power of machine learning and AI for regulators (Nov 2017)
- Using AI to keep criminal funds out of the financial system (Dec 2017)
- The future of regulation: AI for consumer good (Jul 2019)
- AI: Moving from fear to trust (Nov 2022)
- Building better foundations in AI (Jan 2023)
- Shaping the rules for a data driven future (Jun 2022)
- Rolling regulation forwards (Nov 2022)
- Support for automated models (Feb 2023)
- **Innovation, AI & the future of financial regulation (Apr 2023) ("ChatGPT")**
- Global AML and Financial Crime TechSprint (May 2018)
- FCA tackling scams faster as part of data strategy (Jun 2022)
- Asset management: a regulatory perspective (Mar 2018)
- Drivers of change in the financial services industry and how we are responding (Nov 2021)
- AI and financial crime: silver bullet or red herring? (Nov 2018)
- Artificial Intelligence in the boardroom 2019
- **"Maximising the benefits of Generative AI for the digital economy" (19 July 2023, DRCF)**

UK Health Security Agency (UKHSA)

Website

5.7.23

- Artificial intelligence

- Covid 19 – LFD digital reader evaluation (Nov 2022)
- Innovative new ways to improve sexual health and HIV outcomes

OFWAT

Website

5.7.23

- Artificial intelligence
- Frontier model

- OFWAT – Innovation and efficiency gains from the totex and outcomes framework (Jun 2018)
- Europe Economics – Additional evidence relating to frontier shift (Oct 2020)
- PR24 and beyond: Final guidance on long-term delivery strategies (Apr 2022)
- Open data in the water industry: a case for change (Oct 2021)
- Water breakthrough challenge 1 decision document (Sep 2021)
- Reasons for waiving Tideway's procurement obligations (Sep 2020)
- Ofwat's response to consultation on encouraging innovation in regulated utilities (Jan 2019)
- Operational resilience discussion paper (Apr 2022)
- A water supply-demand model for Eng and Wales (Univ of Manchester) (Jul 2021)
- Innovation in Water Challenge Round 1 Decision Document (May 2021)
- Frontier Economics report – Economic impacts of Covid 19 on the water sector (Dec 2020)
- Hydrogen-power from sewage and clean river tech win share of Ofwat fund (Sep 2021)
- Ofwat's Innovation Fund launches £4m open competition (Jan 2023)
- 2021 Water Breakthrough Challenge (May 2021)

OFGEM

Website

5.7.23

- Artificial intelligence

- Case study (UK): Electric vehicle-to-grid (V2G) charging (Jul 2021)
- Ofgem Innovation Link case studies (Jun 2019)
- Ofgem Innovation Vision 2021 - 2025
- Consultation NGET Non-operational IT Capex for addition AI Capabilities (Sep 2021)
- Robotic Roadworks and Excavation System
- Digital Strategy Consultation 2020
- Decision on the Strategic Innovation Fund: round 1 Discovery Phase (unsuccessful projects redacted) (Mar 2022)
- Review of GB energy system operation (Jan 2021)
- Response to questions from the National Data Strategy (Dec 2020)
- ESO roles Guidance (Mar 2023)
- Consultation on frameworks for future systems and network regulation: enabling an energy system for the future (Mar 2023)
- Expert Assessors' Recommendations Report re Strategic Innovation Fund Round 1 2021

Blogs:

- Switching: the future is ... bot? (2017)
- We need a revolution in how and when we use energy (Jul 2021)

Office for Nuclear Regulation

Website

5.7.23

- Artificial intelligence

- External expert panel – regulation of artificial intelligence in nuclear
- The impact of AI/ML on Nuclear Regulation (Jun 2021)
- ONR's approach to regulating innovation
- Office for Nuclear Regulation Regulatory Research Strategy 2019
- Research supporting regulatory guidance for new technologies and new materials
- ONR-RRR-133 (Research into technical and regulatory approaches that enable the safe and secure deployment of AI) delayed.
- Chief Nuclear Inspector's annual report on great Britain's nuclear industry Oct 2022

News:

- Looking to the future: AI in nuclear (Mar 2022)
- Innovation workshop on AI and data science in nuclear (Jun 2022)
- ONR leads expert discussion on AI in nuclear industry (Aug 2022)
- ONR successful in Regulators' Pioneer Fund application (Nov 2022)
- ONR attends conference focusing on Navigating Nuclear Future (Mar 2023)
- Expert panel sessions explore use of AI in nuclear sector (Jun 2023)

Bank of England Prudential Regulation Authority (PRA) Website 5.7.23 • Artificial intelligence • Artificial General Intelligence • Frontier Models	Publications: • Software validation and AI in finance – a primer (Oct 2021) • **Comparing minds and machines: implications for financial stability (Aug 2021) ("artificial general intelligence")** • Delivering data standards and transforming data collection in financial services (Apr 2021) • Transforming data collection from the UK financial sector: a plan for 2021 and beyond (Feb 2021) • Uncertainty and economic activity: a multi country perspective (Jun 2018) ("frontier models")158 Blogs: • Should economists be more concerned about Artificial Intelligence? (Mar 2017) • Opening the machine learning black box (May 2019) • New machines for the old lady (Nov 2017) • Using machine learning to understand the mix of jobs in the economy in real-time (Aug 2018) • Tell me why! Looking under the bonnet of machine learning models (Nov 2019) • Will a robot takeover my job? (Mar 2019) • 2022 press release – BIS Innovation hub Prudential regulation: • DP5/22 - Artificial Intelligence and Machine Learning • SS1/23 – Model risk management principles for banks • CP6/22 – Model risk management principles for banks • Overview of the PRA strategy • PRA Practitioner Panel and Insurance Sub-committee – Annual Report 2020/21 • Prudential Regulation Authority Business Plan 2022/23 • See 2023/24 Research: • The AI Public-Private Forum: Final report • MindBridge Analytics Inc proof of concept (Jul 2017) • Facilitating firms' use of technology, like the cloud, to increase their operational resilience • Call for papers 2023
Pensions Regulator Website 5.7.23 • Artificial intelligence	• FOI request (Jan 2022) • Regulating the pensions and retirement income sector: our joint regulatory strategy (Oct 2018) •

---

158   This paper refers to "frontier models" but that phrase is used in a quite different context to artificial intelligence.

General Medical Council (GMC)

Website

5.7.23

- Artificial intelligence

- GMC-response-to-Committee-on-Standards-in-Public-Life-on-Artificial-Intelligence-and-Public (Feb 2020)
- Horizon Scanning – findings from the first cycle report (June 2018)
- GMC's response to CQC's strategy consultation (Mar 2021)
- Business plan 2019
- The Future Operating Environment for Professional Medical Regulation – Final Report (Aug 2016)
- The impact of service change on doctors' training (Mar 2021)
- How to transform UK healthcare environments to support doctors and medical students to care for patients (see Case Study 20)

Council meetings:

- Council Meeting – 24 February 2022 – Approval to consult on revised draft
- Council Meeting – 25 Feb 2021 – Report of the Executive Board 2020
- Council Meeting – 27 April 2023
- [NB number of specialty curriculums ignored I.e. radiology curriculum]
- Clinical radiology specialty training curriculum (Aug 2020)
- See 2021 (no difference)
- Radiology - Specialty specific guidance

Regulator of Social Housing

Website

5.7.23

- No concepts identified

Payment Systems Regulator (PSR)

Website

5.7.23

- Artificial intelligence

- Payments Transaction Data Sharing and Data Analytics – Strategic Solution – Scope and Governance Oversight (Nov 2017)
- The Horizon for Payments - Report from the Horizon Scanning Working Group (July 2016)
- Data in the payments industry (Sep 2019)
- Blueprint for the Future of UK payments (Jul 2017)

ORR (Office of Rail and Road)

Website

5.7.23

- Artificial intelligence

- Annual efficiency and finance assessment of Network Rail 2021 –22 (Oct 2022)
- Innovation Efficiency Study (Mar 13)
- Annual Report of Health and Safety on Britain's Railways (Jul 2022)

Blog:

- Improving the Human: Data interface within ORR (Nov 2020)

The Office for Professional Body Anti-Money Laundering Supervision (OPBAS) Website [NB OPBAS has no separate website to FCA i.e. cannot do a specific search. Resources will be included in the specific search in FCA at No.26.] [This row is used to individual search the documents linked in this web page]. 5.7.23 • Artificial intelligence	• OPBAS Sourcebook (one footnote reference) • OPBAS anti-money laundering report 2019
Financial Reporting Council Website 5.7.23 • Artificial intelligence • ChatGPT	• Call for participants – AI consultation (Mar 2018) • AI and corporate reporting: How does it measure up? (Jan 2019) • Infographic AI (Feb 2019) • Blockchain and the future of corporate reporting: How does it measure up? (Jun 2018) • Call for evidence - FRC research on the use of artificial intelligence (AI) and machine learning (ML) in actuarial modelling in the UK (Jan 2023) • JOINT FORUM ON ACTUARIAL REGULATION RISK PERSPECTIVE: 2018 UPDATE (April 2019) • **Insight report: AI, Emerging Tech and Governance (Mar 2023) ("ChatGPT")**
The Care Inspectorate Website 7.7.23 • Artificial Intelligence	• Enabling, Connecting and Empowering: Care in the Digital Age • Reference to body in Scotland's Artificial Intelligence Strategy, March 2021
Care Council for Wales (CCW) Website 7.7.23 • Artificial Intelligence	• A Health Wales: Our Workforce Strategy for Health and Social Care, October 2020. • A Healthier Wales: Our Plan for Health and Social Care, 2021. • The Parliamentary Review of Health and Social Care in Wales, 2018.
Northern Ireland Social Care Council (NISCC) Website 7.7.23	No results.

Scottish Social Services Council (SSSC) Website 7.7.23 • Artificial Intelligence	• Data Strategy for health and social care consultation – SSSC response, August 2022 • SSSC Digital Strategy 2021-2024
Human Fertilisation and Embryology Authority Website 7.7.23 • Artificial Intelligence	• Annual report and accounts 2021/22
Human Tissue Authority (HTA) Website 7.7.23 • Artificial Intelligence	• HTA Strategy 2020-23
Nursing and Midwifery Council (NMC) Website 7.7.23 • Artificial Intelligence	• Many references, for example, NMC to White Paper, 21 June 2023. • Strategy 2020 - 2025
General Pharmaceutical Council Website 7.7.23 • Artificial Intelligence	• E.g. Council Meeting, 2022
Pharmaceutical Society of Northern Ireland (PSNI) Website 7.7.23	No results.

The Advisory Conciliation and Arbitration Service (ACAS)

Website

7.7.23

• Artificial Intelligence

• Technology in the workplace: research and commentary

The Ada Lovelace Institute's work relating to foundation models

Explainer: 'What is a foundation model?'
adalovelaceinstitute.org/resource/foundation-models-explainer/

Foundation models in the public sector (policy briefing)
adalovelaceinstitute.org/policy-briefing/foundation-models-public-sector/