Policy briefing

October 2023

Foundation models in the public sector

Foundation models are a form of artificial intelligence (AI) system designed to produce a wide variety of outputs. 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), and image generators like MidJourney.

There is some optimism in policy, the public sector and industry about the potential for these models to enhance public services in the context of budgetary restraints and growing user needs. However, there are also risks around issues like biases, privacy breaches, misinformation, security threats, overreliance, workforce harms and unequal access.

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

For more information about the Ada Lovelace Institute or to discuss this policy briefing, contact our policy team: hello@adalovelaceinstitute.org

As AI technologies advance rapidly, Government must consider carefully how to use foundation models in the public sector responsibly and beneficially. This briefing provides policymakers and public-sector leaders with information to support this.

Key considerations for deploying public-sector foundation models

Use of foundation models in government offices is inconsistent. There is evidence of foundation model applications (such as ChatGPT) being used on an informal basis by individual civil servants and local authority staff. Authorised use of foundation models in the public sector is currently limited to demos, prototypes and proofs of concept.

There is some optimism in policy, the public sector and industry about the potential for foundational models to enhance public services in the context of budgetary restraints and growing user needs. Proposed use cases for foundation models in the public sector include automating the 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 (improving public enquiries management) and consolidating knowledge spread across databases into memos (knowledge management).

Effective use of foundation models by public-sector organisations will require them to carefully consider alternatives and counterfactuals. 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 Nolan Principles of Public Life, which include accountability and openness.

Procurement of foundation models for public sector use is likely to be challenging. There are risks in overreliance on private-sector providers, including a potential lack of alignment between applications developed for a wider range of private-sector clients and the needs of the public sector.

Risks associated with foundation models include biases, privacy breaches, misinformation, security threats, overreliance, workforce harms and unequal access. Public-sector organisations need to consider these risks when developing their own foundation models, and should require information about them when procuring and implementing external foundation models.

Improved governance of foundation models in the public sector will be necessary to ensure the delivery of public value and prevent unexpected harms. These could include:

- regularly reviewing and updating guidance to keep pace with technology and strengthen their abilities to oversee new AI capabilities
- setting procurement requirements to ensure that foundation models developed by private companies for the public sector uphold public standards
- requiring that data used for foundation model applications is held locally
- mandating independent third-party audits for all foundation models used in the public sector, whether developed in-house or externally procured
- monitoring foundation model applications on an ongoing basis

- continuing to implement the Algorithmic Transparency Recording Standard[2] across the public sector
- incorporating meaningful public engagement in the governance of foundation models, particularly in public-facing applications
- piloting new use cases before wider rollout in order to identify risks and challenges
- providing training for employees working with (either developing, overseeing or using) foundation models.

What are foundation models?

Foundation models are a form of AI system designed for a wide range of possible applications, with the capability to complete a range of distinct tasks, including translating and summarising text, generating a rough first draft of a report from a set of notes, or responding to a query from a member of the public with text and images.

Foundation models are already being integrated into commonly used applications: Google and Microsoft's Bing embed them into search engines, Photoshop integrates image generation models,[3] and firms like Morgan Stanley use large language models (LLMs) for internal knowledge search and retrieval.[4]

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 serve as the 'building block' of hundreds of AI applications. Many end users will access these systems via existing tools, such as operating systems, browsers, voice assistants and productivity software (for example Microsoft Office and Google Workspace).

Foundation models are distinct from narrow AI systems, which are trained for one specific task and context. Current foundation models are defined by their scale. Developing them or 'training' them requires access to billions of words of text and computing power that can cost millions of pounds.[5] Future foundation models may not necessarily have these properties.[6]

---

2 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.

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

4 '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.

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

6 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.

The capabilities of foundation models are rapidly evolving and new functionalities are being developed, such as tool-assisted foundation models that can access external data sources such as search engines, and 'text-to-action' models that can autonomously carry out tasks.

Different potential trajectories for the future of a foundation model ecosystem, whether towards concentration of power in a few private companies or a proliferation of open-source alternatives, could shape public-sector procurement and policy.

Uses of foundation models in the public sector

There is evidence of foundation model applications (such as ChatGPT) being used on an informal basis by individual civil servants and local authority staff.[7][8] Formal use of foundation models in the public sector is currently limited to demos, prototypes and proofs of concept.

Proposed use cases for foundation models in the public sector include:

- automating the 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 public enquiries management)
- retrieving and organising knowledge spread across public-sector organisations (knowledge management).

It has been claimed that the use of foundation models for these purposes could lead to greater efficiency in public-service delivery, more personalised and accessible presentation of government communication tailored to individual needs, and improvements in government's own internal knowledge management. However, these benefits are unproven and remain speculative.

There is a risk that foundation models are adopted because they are a new technology, rather than because they are the most suitable solution to a problem. Public-sector users should therefore carefully consider the counterfactuals before implementing foundation models.

This means comparing proposed use cases with more mature and tested alternatives that might be more effective, provide better value for money or pose fewer risks – for example, employing a narrow AI system or a human employee to manage public enquiries rather than building a foundation model-powered chatbot. Evaluating these alternatives should be guided by the Nolan Principles of Public Life, which include accountability and openness.

---

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

8 ibid.

Deploying foundation models in the public sector

If foundation models are judged to be the most appropriate option for addressing a particular problem, there are options that public-sector organisations can choose between in order to deploy them.

A public-sector organisation can buy access to foundation models or foundation model-powered tools from private-sector providers, develop its own foundation models or foundation model-powered applications, or it could do a combination of both.[9]

There are risks in overreliance on private-sector providers, including a potential lack of alignment between applications developed for a wider range of private-sector clients and the needs of 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.

Conversely, developing bespoke public-sector foundation models to replicate or compete with foundation models such as GPT-4 is unlikely to unlock significant public value at proportionate cost.[10]

It could therefore be beneficial for public-sector users to act as 'fast followers', adopting established technologies and practices once they have been tried and tested, rather than always trying to use or develop the latest options. This would not necessarily require procuring external tools and could help to ensure that the deployment of foundation models happens steadily, using proven methods.

It could also 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.

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

10 Ada Lovelace Institute, Regulating AI in the UK (2023) https://www.adalovelaceinstitute.org/report/regulating-ai-in-the-uk/ accessed 1 August 2023.

Risks of using foundation models in the public sector

There are numerous risks and potential harms common to any use of algorithmic systems.[11] Researchers at DeepMind have 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 a foundation model, or treating them as if they are human.
- **Automation, access and environmental harms** arising from the environmental or downstream economic impacts of the foundation model.

All these harms should be considered as possible challenges in the present or near-term deployment of foundation models in a public-sector context. Many of these concerns are best dealt with by the upstream providers of foundation models at the training stage (for example, through dataset cleaning, instruction fine-tuning or reinforcement learning from feedback).[12] Public-sector organisations need to consider potential harms when developing their own foundation models, and should require information 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 the process of using data to prompt the underlying large language model does not violate privacy rights by sharing data with a private provider, for example.

---

11 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

12 Laura Weidinger and others, 'Taxonomy of Risks Posed by Language Models', 2022 ACM Conference on Fairness, Accountability, and Transparency (Association for Computing Machinery 2022) 222 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

Governing foundation models in the public sector

There are steps that policymakers could take to support more effective governance of foundation models in the public sector, mitigating risks and ensuring better outcomes. These include:

Regularly reviewing and updating guidance. Guidance is slowly emerging on interpreting UK laws and regulations in the context of AI, including foundation models, but the regulatory environment remains complex and lacks coherence.[13] Regular reviews of the available guidance would motivate regulators to keep pace with technology and strengthen their abilities to oversee new AI capabilities.[14]

Setting procurement requirements to ensure that foundation models developed by private companies for the public sector uphold public standards.[15] To do this, provisions for ethical standards should be introduced early in the procurement process. These standards should be explicitly incorporated into tenders and contractual agreements.

Requiring that data used for foundation model applications is held locally. Some types of sensitive public data cannot be held outside of the UK for legal reasons. For use cases involving these types of data, Government could seek to ensure that cloud providers locate more data centres physically in the UK.

Mandating independent third-party audits for all foundation models used in the public sector, whether developed in-house or externally procured. Audits would enable AI systems to be properly scrutinised, minimising the risks from public-sector use of foundation models. This would create a strong financial incentive for companies to undertake audits, which in turn would stimulate the growth of a robust UK AI auditing ecosystem.[16]

---

13 See, for example: 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..

14 Commissioned AI Law Consultancy analysis of relevant legislation, regulations and guidelines on AI.

15 Committee on Standards in Public Life, 'Artificial Intelligence and Public Standards: A Review by the Committee on Standards in Public Life' (2020) 8 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

16 Ada Lovelace Institute, AI Assurance? Enabling an ecosystem of risk assessment (2023) https://www.adalovelaceinstitute.org/report/risks-ai-systems/ accessed 16 August 2023.

Incorporating meaningful public engagement in the governance of foundation models, particularly in public-facing applications. While public-sector organisations have existing mandates, deploying AI systems raises new questions of benefits, risks and appropriate use.[17] These questions could be addressed with the support of methods such as citizens' assemblies and participatory impact assessments to involve informed members of the public in decisions about proposed uses.

Piloting new use cases before wider rollout in order to identify risks and challenges. These use cases should be easily understandable even by users with low familiarity with technologies.[18] As the public sector becomes more proficient with foundation model uses, it could progress to conducting large-scale pilots and A/B testing. The results of pilots should be shared across the public sector to support improved assessment of potential use cases over time.

Providing training in relevant technical skills for employees working with (either developing, overseeing or using) foundation models. This would ensure that public and private-sector providers of public services can engage proficiently with suppliers and manage risks.[19]

Monitoring foundation model applications on an ongoing basis. Monitoring and evaluation should continue beyond the initial development and procurement of foundation model applications, incorporating assessment of real-life data and feedback from the operation of the applications.[20]

Continuing to implement the Algorithmic Transparency Recording Standard[21] across the public sector. Pioneered by the Centre for Data Ethics and Innovation, this helps public-sector organisations to provide clear information in a standardised way about the algorithmic tools they use, and why they're using them. Rolling this out more widely, and requiring that foundation model uses are incorporated, would provide a more systematic understanding of how widespread foundation model applications are across the public sector.

17 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/>.

18 'Ada Lovelace Institute (n 9).

19 Committee on Standards in Public Life (n 15) 9.

20 'London Office of Technology and Innovation Roundtable on Generative AI in Local Government' (n 7). See also similar recommendations made by in: Committee on Standards in Public Life (n 15) 9.

21 CDDO and CDEI (n 2).

Next steps

Foundation models may offer an opportunity to address certain challenges in public-service delivery, but Government must take coordinated action to develop and deploy them responsibly, safely and ethically.

Our briefing represents an initial mapping of the issues these systems raise in public-sector contexts, but further research and policy development is required in this fast-evolving field. If you would like more information on our work in this area, or if you would like to discuss implementing our recommendations, please contact our policy team at hello@adalovelaceinstitute.org.

Ada's work relating to foundation models

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

Foundation models in the public sector (full evidence review)
www.adalovelaceinstitute.org/evidence-review/foundation-models-public-sector/

The Ada Lovelace Institute (Ada) is an independent research institute with a mission to make data and AI work for people and society. This means making sure that the opportunities, benefits and privileges generated by data and AI are justly and equitably distributed.

Ada Lovelace Institute
100 St John Street, London, WC1B 3JS
+44 (0) 20 7631 0566

Website: adalovelaceinstitute.org
Twitter: @AdaLovelaceInst
Email: hello@adalovelaceinstitute.org

Registered charity 206601