DEMOS

GB CLOUD

BUILDING THE UK'S CLOUD COMPUTE CAPACITY

DAVID LAWRENCE
ELIZABETH SEGER

In partnership with

UKDayOne

MAY 2024

Published by Demos May 2024

15 Whitehall, London, SW1A 2DD
T: 020 3878 3955
hello@demos.co.uk
www.demos.co.uk

CONTENTS

ABOUT THE AUTHORS

David Lawrence

David Lawrence is the co-director of the UK Day One Project. He has worked in Westminster policy for the last decade, including at Chatham House, in Parliament, and as a parliamentary candidate. He holds degrees from the University of Oxford and the London School of Economics.

Elizabeth Seger

Elizabeth Seger is the Director of Technology Policy at Demos. Elizabeth is internationally recognised for her AI governance work on AI democratisation. She holds a PhD in Philosophy of Science and Technology from the University of Cambridge.

ABOUT THE ORGANISATIONS

UKDayOne

The UK Day One Project is a nonpartisan initiative to accelerate growth and progress in the UK, through crowdsourcing ideas from the science, technology and innovation communities, and turning these into implementation-ready policies.

For more information about the UK Day One Project, please contact *david@ukdayone.org* or *julia@ukdayone.org*

DEMOS

Demos is the UK's leading cross-party think tank. We put people at the heart of policy-making by building bridges between citizens, politicians, policy-makers, and experts from industry, civil society and academia to imagine the new ideas, create the policies, and to tell the stories that will transfrom Britain.

ABOUT THIS PAPER

This project spans Demos's strategic focus areas on *'Trustworthy Technology'* and *'Public Service Reform'*. With emerging technologies transforming our world at an ever faster pace, we work to build bridges between politicians, technical experts, and citizens to explore solutions, build trust, and create policy to ensure our technologies benefit the public.

In this policy brief we discuss how access to the computational resources that underpin AI technologies are becoming a critical linchpin in the realisation of AI benefits for the UK public. In particular, there are great hopes that AI can help support the UK's public services by reforming and providing new services and by improving responsiveness and personalization. However, these goals can not be delivered without reliable access to adequate and secure computing resources to train and run the models in use. Large corporate computer providers currently supply most compute resources to the UK, and while the UK government should not and can not aim to compete with these providers, it is important that the UK can supply enough compute to support its own public sector where private and sensitive data must be processed. More so, publicly owned compute will facilitate more seamless and secure integration of AI into public service, closer oversight over an increasingly critical resource, and an opportunity to engage citizens in decision-making about how the resource is used ensure that AI solutions are being developed to address the societal challenges that matter most to the public.

SUMMARY

Artificial Intelligence (AI) is likely to be the most transformative general-purpose technology of our time. Its development and adoption has the potential to accelerate UK growth and prosperity. Compute capacity is essential for training and running AI models. Access to cloud compute could become as important for the UK's economy, society and security as access to the internet, electricity or oil and gas.

The UK Government has already committed £1.5bn to compute infrastructure for academic and industry research through its exascale and AI Research Resource (AIRR) programmes, since March 2023. Additionally, the Government is poised to spend £7.5bn on private contractors (such as AWS, whose revenue from the Government increased 76% last year) to provide cloud compute for public services over the next 3 years.

The UK should build on this with a £1.5 billion investment in publicly owned compute. This could be through a new publicly owned company - GB Cloud.

GB Cloud would:

1. Reduce dependence on foreign hyperscalers and big tech companies.
2. Support public services by providing the UK state with reliable, protected and publicly-owned data centres.
3. Protect data privacy by ensuring sensitive data used in the provision of public services (e.g. public health and DWP data as well as military, and intelligence data) is held and processed domestically and not shared with international, private compute providers.
4. Support UK growth and innovation by providing further cloud computing access to UK AI start-ups, universities and research labs where GB Cloud is in excess.

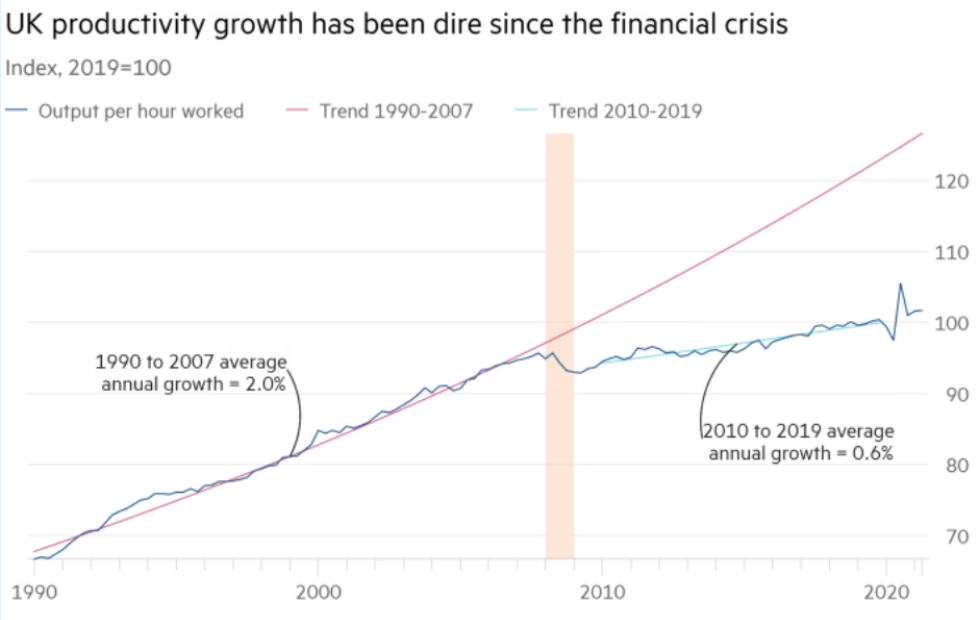
Over the course of the next Parliament, this initial £1.5 billion investment in GB Cloud could be phase-funded up to £10 billion through a combination of private and public investment.

THE CHALLENGE

UK productivity - and therefore real wage growth - has stalled since the financial crisis (figure 1). Brexit, Covid and domestic political instability have further compounded the UK's economic woes. UK public services are struggling, while the cost of living and national debt remain high. The UK has fallen behind many of its peers in terms of growth, productivity and wages.

Meanwhile, the UK faces new security challenges. Russia's invasion of Ukraine highlighted the fragility and interdependence of European supply chains, particularly in energy. US-China tensions, as well as the recent Covid pandemic, reinforce the importance of domestic industrial capacity. New technologies bring new threats, from cyber attacks to biological warfare.

FIGURE 1
UK PRODUCTIVITY GROWTH SINCE THE FINANCIAL CRISIS



Sources: ONS, FT calculations © FT

The new government in 2024 will need to deliver on economic growth while rebuilding public services and defending the country against national security threats. These three goals pose a significant challenge for the new government, especially in the context of fiscal constraints.

THE OPPORTUNITY

ARTIFICIAL INTELLIGENCE (AI) CAN HELP ADDRESS THESE CHALLENGES

A significant part of the UK is already concerned about the risks of drought in the UK. When asked, one in AI is likely to be the most transformative general-purpose technology of our time. Its development and adoption has the potential to alter the UK's growth trajectory. AI is relevant to delivering productivity growth, catalysing scientific progress, helping the UK reach net Zero, discovering new medical breakthroughs, and improving public services and education (See Box 1). To deliver these goals, the UK will need to increase AI use across the public sector.

AI POWERED UK

AI can support UK growth and prosperity through:

- **Boosting productivity** and efficiency across various sectors, including manufacturing, healthcare, finance, and logistics, by automating tasks, optimising processes, and enabling data-driven decision-making. There is significant evidence that existing AI models can already increase productivity, helping workers complete tasks up to 56% faster in some cases. 18% of UK employment is in the public sector. Supporting productivity growth in this area requires state technology adoption.
- **Higher wages for workers and lower costs for consumers** are associated with improved productivity. However the AI rollout needs to be coupled with upskilling opportunities to minimise wage loss due to job loss during the transition.
- **Driving innovation and competitiveness** by enabling the development of new products, services, and business models leveraging AI capabilities such as machine learning, computer vision, and natural language processing.
- **Attracting investments and fostering entrepreneurship** by positioning the UK as a global hub for AI research, development, and talent, encouraging the growth of AI-based startups and the creation of high-skilled jobs.

AI can support public services by:

- **Providing new services:** public services have suffered from austerity and lack of ambition. AI offers the potential for reforming and providing new services in several existing departments, including the NHS, education, transportation, and DWP.
- **Improving interaction and responsiveness:** AI could improve individuals' engagement with public services and the political system through offering faster and more personalised services, and new platforms for citizens to productively express their concerns.

- **Lowering costs:** AI could lead to efficiency gains for government departments and services such as the NHS, through automation, faster scaling of new projects, and more efficient resource allocation.

AI can contribute to Net Zero goals by:

- **Optimising energy grids** through AI-powered predictive analytics and demand forecasting to enable more efficient use of renewable energy sources and to reduce energy waste.
- **Enhancing energy efficiency in buildings** and infrastructure by using AI to monitor and optimise heating, cooling, lighting, and other systems.
- **Reducing traffic emissions** through AI-based traffic management, route optimization, and autonomous vehicle technologies.
- **Modelling environmental factors**, such as greenhouse gas emissions, deforestation, and climate patterns, to inform data-driven policymaking and targeted mitigation.
- **Accelerating the development of new sustainable materials low-carbon technologies** by leveraging AI for simulations, design optimization, and process control.

THE UK NEEDS SOVEREIGN COMPUTE

To safely and reliably deliver on the above goals, the UK must have adequate, secure, and reliable access to the necessary compute resources to power its AI applications.

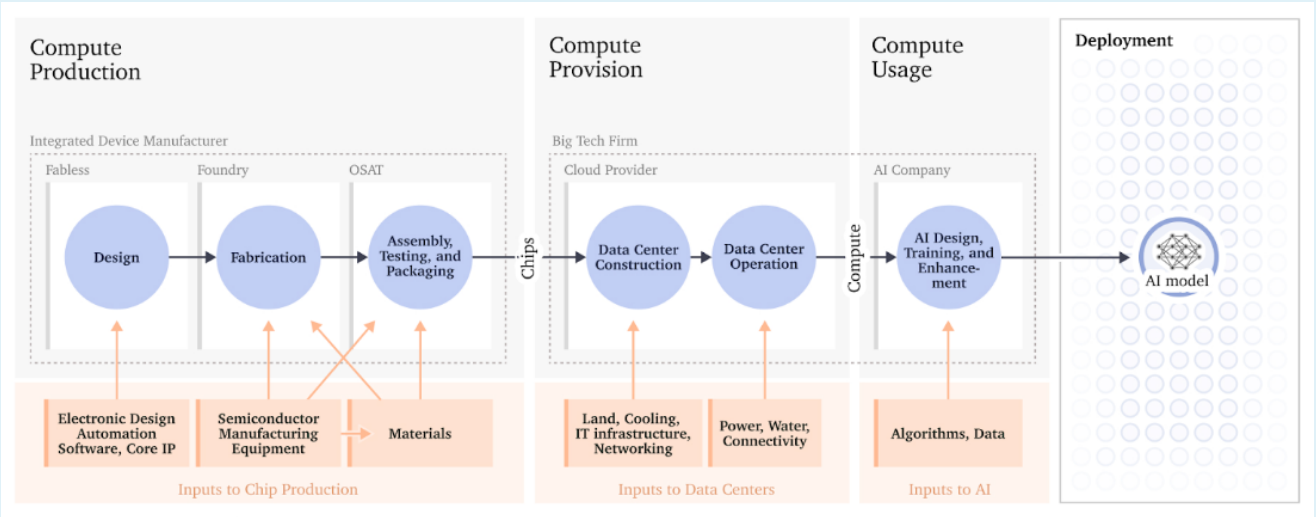
Compute refers to the physical infrastructure and capacity designed to handle intensive computational tasks, enabling the processing and analysis of vast amounts of data at high speeds. Compute is essential for training and running AI models, so much so that all of the big AI start-ups have entered into 'compute partnerships' with Big Tech: e.g. OpenAI and Microsoft, Anthropic and Amazon/Google, Mistral and Microsoft, and DeepMind and Google.

The UK cannot and should not aim to compete with the 'hyperscalers' that provide 60% of global cloud compute: AWS, Microsoft Azure, and Google Cloud (Figure 3). But the UK can provide enough compute to support its researchers, public services, and start-ups so that they are not so heavily reliant on foreign computer providers, and to have a local and secure compute option for processing private and sensitive public sector data.

More so, cloud computing is a good place for the UK to specialise in the AI supply chain (Figure 2). Some parts of the supply chain, such as chip fabrication or making the machines that fabricate chips, are not good fits for the UK; they are near-monopolies (ASML in the Netherlands, NVIDIA in the US and TSMC in Taiwan) and require highly specialised talent and hugely expensive equipment. The UK cannot seriously compete with these companies (Figure 3).

However, thanks to companies like DeepMind, Graphcore, Faculty, Wayve, Healx and many others, as well as several world-leading universities, the UK has a comparative advantage in AI design, training and enhancement (and to some extent in chip design), and this advantage crucially relies on data centre cloud computing.

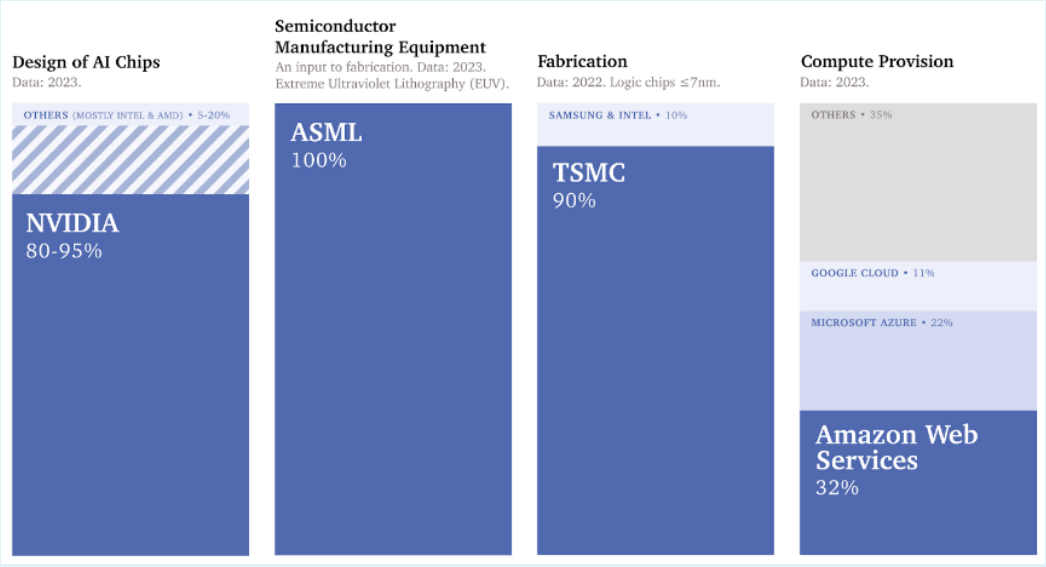
FIGURE 2
OVERVIEW OF THE AI COMPUTE SUPPLY CHAIN



Source: Sastry, Heim, Anderljung, Brundage, Hazell, O'Keefe, Hadfeild et al., (2024)

An overview of the AI compute supply chain. First, chips are produced through a process of design, fabrication, and testing. They are then distributed and accumulated in data centres. Compute users — such as AI developers — can then train and run AI systems from these AI supercomputers.

FIGURE 3
THE SUPPLY CHAIN FOR AI CHIPS IS HIGHLY CONCENTRATED



Source: Sastry, Heim, Anderljung, Brundage, Hazell, O'Keefe, Hadfeild et al., (2024)

Several critical steps-including AI chip design and production-have fewer than three suppliers. These facts highlight how difficult it is to compete at the cutting edge of chip production, and how dependent countries are on a tiny number of companies.

It is not necessary for the UK's AI industry to rely ***exclusively*** on UK based compute resources, but there are still substantial benefits to having sizable sovereign compute capacity on UK soil. It is similar to how we are not self-sufficient in energy or food, but we invest enough in our own supply to protect our people in crises, to reduce dependence and vendor lock-in, and to deter price-gouging and exploitation.

Sovereign compute capacity is important for several reasons:

1. To ensure reliable compute access

Reliable access to compute resources becomes increasingly critical as AI becomes more deeply integrated into every aspect of our lives. If in the future AI comes to underpin our economies, public services, communications, and social interaction, then losing access to reliable computers could be catastrophic, akin to the impact losing national broadband or electrical infrastructures would have today. It is a matter of economic and national security.

We should also not underestimate the potentiality of such an occurrence. Unlike some other digital technologies, AI relies heavily on physical infrastructure, from data centres to the energy required to train models, as well as the most advanced semiconductors. Relying on foreign compute resources means relying on foreign governments to keep their data centres safe and secure and to share those resources internationally. Meanwhile, compute capacity is quickly becoming a geopolitical bargaining tool and states are increasingly pursuing their own sovereign cloud compute to defend national security. Cloud computing has been cited by the US's *Center for Strategic and International Studies (CSIS)* as a key site of geopolitical tension between the US and China and *China has sought various ways to skirt the US's export controls* for computing power, including aiming to boost its own compute capacity. Meanwhile, 75% of countries - particularly in Europe - have introduced *data localisation* laws, requiring data centres on national soil.

2. To maintain reasonable compute costs

Ensuring ample availability of sovereign compute in the UK may help maintain stable and reasonable compute prices, by providing a suitable and accessible alternative for UK AI development. More so, a condition of publicly funding new compute resources could be to prioritise UK AI research institutions and start-up industry with the provisions of reduced or at cost compute. Low compute costs would increase the UK's attractiveness to new AI startups looking for a place to set up.

There are other options for lowering compute costs for UK researchers and AI developers. For instance, the government could offer larger tax breaks for data centre builders and operators, or the government could buy compute credits in bulk from hyperscalers and then offer them to UK startups and researchers at a subsidised rate. However, such subsidies would be a very large and likely inconsistent cost; as compute costs fluctuate, the effective subsidy price would change. But perhaps more importantly, subsidising compute costs in this way further reinforces the dominance of foreign hyperscalers in the UK market

3. For regulation and safety

AI poses several risks, ranging from near-term social harms to longer run impacts on society, the economy and humanity. Sovereign compute infrastructure would be subject to UK regulation that aims to help mitigate AI harms. For example, the UK government could enforce 'know your customer' requirements for UK based data centres, requiring that for large training runs, only known and vetted customers are granted access. Data centres could also be required to restrict the use of sensitive personal data for training.

4. For data protection and security

Where personal data (e.g. health data for medical applications) or classified data (e.g. for use in some chemical or biological research applications) is used to train models, sovereign compute infrastructure ensures that data need not be transferred overseas. Personal data privacy concerns are a driving force behind the EU's own data localisation laws.

5. To make use of difficult-to-tap renewable energy sources

Data centres are very energy intensive to run, but they are great candidates for tapping into the UK's renewable energy potential. Data centres can be located in remote locations where the practicality of building renewable energy infrastructure is limited by the cost of transmitting energy to far off population centres. For example, one option could be to build data centres in Scotland. Scotland offers abundant renewable wind energy potential with tens of megawatts of potential capacity, yet transmission costs to population centres further south limit its utilisation. Building data centres in Scotland could leverage this underutilised renewable capacity by locating near the source.

ADVANTAGES OF PUBLIC OWNERSHIP

AI systems will be used across our economy and society, in almost every company and public service. On-going, secure, reasonably priced access to cloud compute is likely to become as important for the UK's economy, society and security as access to the internet, electricity or oil and gas. Like other public utilities, there are significant benefits to public ownership of - not just public investment in - sovereign cloud compute.

The benefits of public ownership include:

1. A more secure and seamless model for public service delivery

Publicly owned compute infrastructure dedicated to AI could improve the delivery of public services, offering increased security, efficiency, and seamless integration.

Heightened Security:

By hosting AI models and applications on publicly owned, government controlled computing resources, sensitive data and critical decision-making processes can be kept within a secure environment, mitigating the risks of data breaches and unauthorised access. Publicly owned compute is particularly important for minimising the vulnerability of our most sensitive data – such as national security, military, and intelligence data as well as public health and DWP data – that must be used to train models for subject specific use and that must be processed in the day to day operation of AI for public service delivery.

More efficient integration:

Furthermore, a centralised AI compute infrastructure can streamline operations and enhance efficiency across various government agencies and departments. AI models can be trained and deployed more efficiently, leveraging shared resources and eliminating redundancies.

One option is to train a UK Public AI model - such as 'BritGPT' - to serve a common resource to integrate into public service applications. The model would be trained on known and vetted datasets to minimise bias, reflect British values, and to underpin public trust in the use of AI in public services. The French government has created one such sovereign model, Albert, hosted on French public compute infrastructure.

Released in April 2024, Albert has been hailed as "a major turning point in administrative simplification, placing people at the heart of public services." It has realised and projected applications in human resources, tax advising, conducting audits, environmental planning, and throughout legal and medical systems.

As noted above, the UK Government is also about to spend £7.5bn over the next 3 years on private contractors to provide cloud compute to operate AI for public services. This level of spending is baked in - the question is whether all that money should go toward US hyperscalers, or whether we should invest some portion in our own public compute infrastructure with the added benefits of security and efficiency.

2. Closer oversight for an increasingly critical resource

Sovereign compute would be subject to UK regulation, but there is a degree of separation between the provider and the regulator. Government owned and operated public resources can be more tightly controlled and closely monitored. The capacity for direct oversight may become more desirable as AI becomes more prevalent and compute resources grow in importance as a public utility akin to broadband and water. Establishing a public compute option will also make it easier to integrate previously independent infrastructure into public ownership at a later date should that option become desirable.

3. Building public trust and support

British-owned, democratically accountable compute could help underpin public trust in AI, fostering wider adoption of the transformative technology. Engaging citizens in the decision-making processes about the compute resource — for instance, how public compute resources are allocated (e.g. to which public services, to industry, or research) and where data centres are built — can ensure that AI solutions are being developed to address societal challenges that matter most to the public and to improve the lives of citizens. By giving citizens an economic stake and a degree of control over AI development and deployment, the government can create a sense of collective ownership and accountability, mitigating concerns about private interests dominating this critical, social domain.

PLAN OF ACTION

The inputs required for the UK to develop compute capacity are:

- Land - space for the data centres
- Reliable renewable energy - to power the data centres.
- Chips - to populate the data centres
- Skilled workforce - to build and operate the data centres
- Money - to pay for the above

In the first 100 days of government, the Secretary of State for Science, Innovation and Technology should:

1. Set up GB Cloud as a publicly owned company, with a founding constitution to provide publicly-owned computing power to train and run AI models for:
    a. The delivery of public services
    b. UK startups and innovators
    c. UK universities and researchers
2. Through DSIT, or another appropriate body, begin a consultative process for scoping out:
    a. The location of data centres
    b. Public support and benefit from data centres
    c. Areas of economic need or opportunity that could benefit from public cloud
3. Engage multi-stakeholder and citizen consultation to determine how best to set up the infrastructure in terms of governance and ownership to best serve and respond to public needs
4. Commission a team within DSIT to identify potential sites and university partners for building new publicly owned supercomputing data centres. Engage with civil society to consult with local communities about data centre impacts and construction.

Over the next 2 years, the Government should:

5. Begin a procurement process for private sector contractors to build relevant GB Cloud infrastructure and for purchasing chips
6. Do recurring public engagement on resource allocation
7. Open up cloud compute to:
    a. Government digital service
    b. NHS / public services
    c. UK startups
    d. Universities / research labs

BUDGET

We propose an initial investment of £1.5bn from the Government for GB Cloud. This should be spent on the initial consultation phases and procurement processes, and should be enough to cover the construction of 2-3 large data centres and ongoing maintenance and running costs for the first 2 years.

The Government should use this initial injection to 'crowd-in' private, university and philanthropic funding where possible. It could also be used in conjunction with new UKRI grants (e.g. for universities and UK start ups) to support compute infrastructure, where appropriate.

The new Government in 2024 will have limited fiscal headroom and many competing priorities for funding - from public services to wider infrastructure. Spending an additional £1.5bn on public compute is not a small figure, however it is a worthwhile investment in an increasingly essential technology.

The Government has already announced £900m towards a new 'exascale' supercomputer and a dedicated AI Research Resource. As part of this, £225m has been allocated to build the UK's most powerful AI supercomputer with the University of Bristol and HP Enterprise, and £500m will be awarded through UKRI, towards wider cloud compute capacity. Together this totals a £1.5bn investment over the past year.

A further £1.5bn investment in GB Cloud would complement these investments, helping to cement the UK's position as Europe's leading AI hub, while also providing something that the UK does not yet have - publicly owned compute to support public sector AI.

We estimate a rough cost breakdown of:

- **£500m** for an AI supercomputer in addition to those being built in Bristol and Edinburgh to resource cutting-edge AI research and large model training needs within government (e.g. AISI), academia and industry.
- **£1bn** public tender to build two new data centres to support the development, deployment, and day-to-day utilisation of AI for public services and to support the data centres' running and maintenance for 2 years.

Over the next Parliament (5 years), the Government should aim to repeat this process annually, drawing on any lessons learned, and spending up to £10bn on GB Cloud. During this period, the Government should aim to reduce its current spending on private cloud compute, and replace this with funding GB Cloud, which in turn can support the Government's computing needs and public services.

This £10bn figure is a guideline and ambition. If the project is going well and/or the price of private compute increases, the Government may want to consider investing more in GB Cloud. Similarly, if the project returns less value for money than expected, GB Cloud should focus on providing a secure, public fallback option, while private compute can continue to be the primary support for the Government's computing needs and public services.

Another key test for scaling up the project is the level of demand from UK start ups, universities and researchers. Access to compute will be essential for AI research and innovation in the UK. DSIT should review the extent to which researchers and academics demand GB Cloud's services, and how it compares to private options and compute accessibility for AI research and development through UKRI grants.

Finally, the Treasury, DSIT or other appropriate body should make an assessment of GB Cloud's ability to 'crowd in' private and philanthropic funding. Universities, for instance, may be willing to invest part of their own budgets on GB Cloud projects, either for public benefit or in order to support their researchers.

Although £10 billion represents a significant increase in the Government's existing spending on UK compute infrastructure, it remains lower than other major infrastructure projects. For comparison, Crossrail (the Elizabeth Line) cost £20 billion. The Sizewell C nuclear power plant is likely to cost £35 billion. The new Dreadnaught-class Trident submarines are likely to cost £40 billion. The HS2 project is estimated at around £100 billion.

Compared to other AI investments: building a new chip factory or 'fab' costs £1-10 billion. Foundation models cost £10-100 million to train per model just in compute costs (including energy and infrastructure, but excluding salaries of talented staff), and in coming years this may increase to the £1-10 billion range; building additional compute infrastructure in the UK to support in the development and running of public sector AI will ensure more of this spending stays in the UK.

FAQs

1. Who are the main cloud compute providers for the UK?

The 'Big Three' cloud providers are Amazon Web Services (AWS), Microsoft Azure and Google Cloud. The trainers of big, generally capable and adaptable 'foundation models' – notably OpenAI, DeepMind, Anthropic, Stability AI –- have mostly all entered into 'compute partnerships' with the Big Three. There is nothing wrong with compute provision via partnership per se, but all these providers are US-based, and the US's compute dominance helps it to attract and nurture the most advanced AI companies. For example, DeepMind is a UK-based company, but it is owned by Google and relies on Google for its Cloud Compute. *Microsoft has also committed £2.5 billion* to building data centre infrastructure in the UK which will double the UK's current data centre footprint - essential for the deployment, serving, and day-to-day utilisation of foundation models. While this investment is certainly welcome, GB Cloud could help reduce reliance on US hyperscalers and offer preferential access to UK research labs and start-ups, helping to develop a home grown AI ecosystem.

In 2012, the UK set up its G-Cloud procurement to make it easier for British cloud providers to win public sector contracts. However, this was reversed by the Government in 2016, and since then the Big Three has dominated the UK market, making it harder for SME cloud providers.

2. What is the current state of UK investment in sovereign compute, and where are we coming up short?

The UK has recognized the importance of investing in sovereign compute capacity. Since the publication of the independent *Future of Compute Review* (March 2023), the government has announced investments over £1.5 billion across the AI Research Resource (AIRR) and the Exascale Project.

Two new supercomputers will soon be operational as part of the AIRR network, Dawn (University of Cambridge), and Isambard-AI (University of Bristol), and the first *Exascale system* will be hosted by the University of Edinburgh.

UKRI also invested £210 million in partnership with IBM to establish the Hartree National Centre for Digital Innovation which has so far announced the construction of a new *data centre* and *supercomputer* to support AI and quantum computing research. Recently, the Hartree National Centre for Digital Innovation (HNCDI) programme (HNCDI) announced a complimentary *$38 million* funding commitment towards a new supercomputing centre.

While this investment in sovereign compute infrastructure has been significant, we are making up for missing past investment rounds and are now working to catch up with other nations. The *UK ranks tenth* globally in supercomputer capacity coming in behind Italy, Finland, and Russia.

More so, the current UKRI funded AI compute resources are primarily targeted for use in cutting-edge academic and industry research while we still rely heavily on big three compute providers to support the development and use of AI by the public sector and SMEs.

3. What are others doing in this space?

- In the US, *the Biden administration has allocated $25 billion* to emerging technologies. The US Defense Information Systems Agency *listed cloud computing* as one of three strategic priorities, given the centrality of cloud compute to national security, and recently spent $4.4 billion on its defence cloud contract. The US's *National AI Research Resource (NAIRR)* has a 6-year plan to spend $750million every 2 years (for a total of $2.6billion over 6 years) for growing the national academic AI research capacity.
- In France, a *Government AI Commission has called for €27 billion* in public and private AI investment over the next five years. This is part of its ambition to make France "major pole of computing power". The report called for France to: "secure collective supply on a national and European scale, solicit projects to set up computing centres in France with a public guarantee of usage and simplification of procedures, [and] set up a tax credit for model training on national computing power".
- In Canada, the Government announced a *$1.46 billion investment* in "computing capabilities" through the AI Compute Access Fund, in 2024.
- In Australia, the government has committed *$36m* to the National AI Centre and four AI and Digital Capability Centres over the next four years. The digital capability centres have a *focus on industry access* to AI compute and services. Australia has its own *National Research Infrastructure* (similar to NAIRR) as well as its own sovereign cloud provider *AUCloud*. AUCloud only sells high security sovereign cloud services to the Australian Government, Defence, Intelligence, and Critical National Industry (CNI) communities and is operated by security cleared Australian personnel.
- In Japan, the government announced *$470m* in funding for a new supercomputer in April 2024. It had already spent *$1.25bn* on its Fugaku supercomputer, completed in March 2021.
- In Singapore, state compute investments are part of a larger 5-year *$743m* investment in AI.
- In the UAE, a *$100m deal* for *Cerebras Systems'* supercomputers was announced in 2023. $3.84m was used to train Falcon (UAE's proprietary model) with *384 A100 chips* worth around *$10,000 each*.
- In Saudi Arabia, the state has purchased a minimum of *3,000 Nvidia's H100 chips* each valued at *$40,000 per processor*. The investment comes to a total of > >$120m for H100s. The Shaheen III supercomputer, uses *2,800 NVIDIA Grace Hopper Superchips*. Each chip is priced at *$65,400*, for a total of $183 million

	COMPUTE INVESTMENT (USD)	
	From Government	From Microsoft (for comparison)
UK	~$1.9bn (*AIRR + exascal* + *HNCDI*)	*$3.16bn*
USA	*$0.75 every 2 years, aiming for $2.6bn over 6 years*	*>$10bn*
Germany	*$0.54bn* for AI (2024) *including supercomputing*	*$3.5bn*
France	*$2bn* (and $10bn called for over next 5 years)	*$4bn*
Canada	*$1.46bn*	*$0.5bn* announced in November 2023

	COMPUTE INVESTMENT (USD)	
	From Government	From Microsoft (for comparison)
South East Asia	Singapore *~$0.743bn*	*$2.2bn to Malaysia and $1.7bn in Indonesia*, a new data centre in *Thailand*
UAE	*>$100m*	*$1.5bn*
Saudi Arabia	*>$0.3bn*	*$2bn* for a cloud data centre
Japan	*$0.47bn* announced in 2024	*$2.9bn*
Australia	*$32.6m* over the next 4 years as of 2022	*$3.2bn* for hyperscal cloud compute and AI infrastructure
TOTAL	**~$10 billion**	**$50-100bn**

4. How might the public respond to GB Cloud?

UK Government polling suggests there is public interest in AI but concerns about its impact, and a desire for Governments to regulate and test models. GB Cloud could provide a way for the UK to invest in AI while building public trust in a new and often misunderstood technology. Since GB Cloud would be publicly-owned and democratically accountable, this may assuage some people's fears that AI is dominated by the interests of a handful of non-UK technology companies.

However, the Government would have to communicate the economic and social benefits of AI – and its strategic importance for security – in order to justify spending public money on cloud compute. The public may hold the Government to a higher standard than technology companies, in terms of data privacy, misuse and AI safety.

As part of embarking on the GB Cloud plan we recommend engaging civil society to conduct extensive multi-stakeholder engagement to inform decisions, for instance, about how the resource is managed or allocated (to which public services, industries, or research areas). Building robust public trust and support will be achieved by putting people at the heart of the public's AI effort.

5. Should we be concerned about the environmental impact of GB Cloud?

Yes. Supercomputers and data centres use a lot of energy and produce a lot of heat. This makes it even more important for the UK to transition towards renewable sources of energy. GB Cloud could lead the world by guaranteeing that all its energy will come from zero carbon sources which would also provide a customer driving consistent demand for renewables.

Net Zero data centre operation is a feasible goal, already committed to by 100 data centre operators and trade associations to be achieved by 2030 as part of the *Climate Neutral Data Centre Pact*. Toward that end, *AWS* has, for example, purchased a nuclear power plant to directly power its adjacent data centre. The UK has abundant wind and tide options, in addition to nuclear, with the option to build data centres away from population centres to be nearer the renewable source.

One option, for example, could be to build data centres in Scotland. Scotland offers abundant renewable wind energy potential with tens of megawatts of potential capacity, yet transmission costs to population centres further south limit its utilisation. Building data centres in Scotland could leverage this underutilised renewable capacity by locating near the source.

6. What accountability and evaluation measures could be included to ensure that the proposal is carried out well?

As with other publicly owned companies, the Government should set out a founding constitution for GB Cloud which ensures that it serves the public benefit. This could include provisions such as:

- Incorporating as a public benefits company
- Ensuring free or discounted access to cloud compute for Government departments, public bodies (e.g. the NHS) and educational institutions
- Ensuring free or discounted access for UK-based startups, scientists, and AI researchers
- Articulating and maintaining high standards for data quality, provenance, and privacy
- Implementing measures to ensure public cloud is not used for training harmful AI applications - such as for political or sexual deepfakes
- Implementing democratic safeguards, e.g. through parliamentary oversight and public consultation
- Establishing a governance structure to facilitate compliance and representation of public interest, e.g. through an independent, democratically elected board to make decisions on behalf of the company to ensure compliance.
- Including provisions for periodically engaging diverse stakeholder and citizen groups in deliberative processes to maintain alignment of UK AI efforts with public interest.

Licence to publish

Demos – Licence to Publish

The work (as defined below) is provided under the terms of this licence ('licence'). The work is protected by copyright and/or other applicable law. Any use of the work other than as authorized under this licence is prohibited. By exercising any rights to the work provided here, you accept and agree to be bound by the terms of this licence. Demos grants you the rights contained here in consideration of your acceptance of such terms and conditions.

1 Definitions

a 'Collective Work' means a work, such as a periodical issue, anthology or encyclopedia, in which the Work in its entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole. A work that constitutes a Collective Work will not be considered a Derivative Work (as defined below) for the purposes of this Licence.

b 'Derivative Work' means a work based upon the Work or upon the Work and other pre-existing works, such as a musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Work may be recast, transformed, or adapted, except that a work that constitutes a Collective Work or a translation from English into another language will not be considered a Derivative Work for the purpose of this Licence.

c 'Licensor' means the individual or entity that offers the Work under the terms of this Licence.

d 'Original Author' means the individual or entity who created the Work.

e 'Work' means the copyrightable work of authorship offered under the terms of this Licence.

f 'You' means an individual or entity exercising rights under this Licence who has not previously violated the terms of this Licence with respect to the Work, or who has received express permission from Demos to exercise rights under this Licence despite a previous violation.

2 Fair Use Rights

Nothing in this licence is intended to reduce, limit, or restrict any rights arising from fair use, first sale or other limitations on the exclusive rights of the copyright owner under copyright law or other applicable laws.

3 Licence Grant

Subject to the terms and conditions of this Licence, Licensor hereby grants You a worldwide, royalty-free, non-exclusive, perpetual (for the duration of the applicable copyright) licence to exercise the rights in the Work as stated below:

a to reproduce the Work, to incorporate the Work into one or more Collective Works, and to reproduce the Work as incorporated in the Collective Works;

b to distribute copies or phono-records of, display publicly, perform publicly, and perform publicly by means of a digital audio transmission the Work including as incorporated in Collective Works; The above rights may be exercised in all media and formats whether now known or hereafter devised. The above rights include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. All rights not expressly granted by Licensor are hereby reserved.

4 Restrictions

The licence granted in Section 3 above is expressly made subject to and limited by the following restrictions:

a You may distribute, publicly display, publicly perform, or publicly digitally perform the Work only under the terms of this Licence, and You must include a copy of, or the Uniform Resource Identifier for, this Licence with every copy or phono-record of the Work You distribute, publicly display, publicly perform, or publicly digitally perform. You may not offer or impose any terms on the Work that alter or restrict the terms of this Licence or the recipients' exercise of the rights granted hereunder. You may not sublicence the Work. You must keep intact all notices that refer to this Licence and to the disclaimer of warranties. You may not distribute, publicly display, publicly perform, or publicly digitally perform the Work with any technological measures that control access or use of the Work in a manner inconsistent with the terms of this Licence Agreement. The above applies to the Work as incorporated in a Collective Work, but this does not require the Collective Work apart from the Work itself to be made subject to the terms of this Licence. If You create a Collective Work, upon notice from any Licensor You must, to the extent practicable, remove from the Collective Work any reference to such Licensor or the Original Author, as requested.

b You may not exercise any of the rights granted to You in Section 3 above in any manner that is primarily intended

for or directed toward commercial advantage or private monetary compensation. The exchange of the Work for other copyrighted works by means of digital file sharing or otherwise shall not be considered to be intended for or directed toward commercial advantage or private monetary compensation, provided there is no payment of any monetary compensation in connection with the exchange of copyrighted works.

c If you distribute, publicly display, publicly perform, or publicly digitally perform the Work or any Collective Works, you must keep intact all copyright notices for the Work and give the Original Author credit reasonable to the medium or means You are utilizing by conveying the name (or pseudonym if applicable) of the Original Author if supplied; the title of the Work if supplied. Such credit may be implemented in any reasonable manner; provided, however, that in the case of a Collective Work, at a minimum such credit will appear where any other comparable authorship credit appears and in a manner at least as prominent as such other comparable authorship credit.

5 Representations, Warranties and Disclaimer

a By offering the Work for public release under this Licence, Licensor represents and warrants that, to the best of Licensor's knowledge after reasonable inquiry:

i Licensor has secured all rights in the Work necessary to grant the licence rights hereunder and to permit the lawful exercise of the rights granted hereunder without You having any obligation to pay any royalties, compulsory licence fees, residuals or any other payments;

ii The Work does not infringe the copyright, trademark, publicity rights, common law rights or any other right of any third party or constitute defamation, invasion of privacy or other tortious injury to any third party.

b Except as expressly stated in this licence or otherwise agreed in writing or required by applicable law, the work is licenced on an 'as is' basis, without warranties of any kind, either express or implied including, without limitation, any warranties regarding the contents or accuracy of the work.

6 Limitation on Liability

Except to the extent required by applicable law, and except for damages arising from liability to a third party resulting from breach of the warranties in section 5, in no event will licensor be liable to you on any legal theory for any special, incidental, consequential, punitive or exemplary damages arising out of this licence or the use of the work, even if licensor has been advised of the possibility of such damages.

7 Termination

a This Licence and the rights granted hereunder will terminate automatically upon any breach by You of the terms of this Licence. Individuals or entities who have received Collective Works from You under this Licence, however, will not have their licences terminated provided such individuals or entities remain in full compliance with those licences. Sections 1, 2, 5, 6, 7, and 8 will survive any termination of this Licence.

b Subject to the above terms and conditions, the licence granted here is perpetual (for the duration of the applicable copyright in the Work). Notwithstanding the above, Licensor reserves the right to release the Work under different licence terms or to stop distributing the Work at any time; provided, however that any such election will not serve to withdraw this Licence (or any other licence that has been, or is required to be, granted under the terms of this Licence), and this Licence will continue in full force and effect unless terminated as stated above.

8 Miscellaneous

a Each time You distribute or publicly digitally perform the Work or a Collective Work, Demos offers to the recipient a licence to the Work on the same terms and conditions as the licence granted to You under this Licence.

b If any provision of this Licence is invalid or unenforceable under applicable law, it shall not affect the validity or enforceability of the remainder of the terms of this Licence, and without further action by the parties to this agreement, such provision shall be reformed to the minimum extent necessary to make such provision valid and enforceable.

c No term or provision of this Licence shall be deemed waived and no breach consented to unless such waiver or consent shall be in writing and signed by the party to be charged with such waiver or consent.

d This Licence constitutes the entire agreement between the parties with respect to the Work licenced here. There are no understandings, agreements or representations with respect to the Work not specified here. Licensor shall not be bound by any additional provisions that may appear in any communication from You. This Licence may not be modified without the mutual written agreement of Demos and You.

DEMOS

Demos is a champion of people, ideas and democracy. We bring people together. We bridge divides. We listen and we understand. We are practical about the problems we face, but endlessly optimistic and ambitious about our capacity, together, to overcome them.

At a crossroads in Britain's history, we need ideas for renewal, reconnection and the restoration of hope. Challenges from populism to climate change remain unsolved, and a technological revolution dawns, but the centre of politics has been intellectually paralysed. Demos will change that. We can counter the impossible promises of the political extremes, and challenge despair – by bringing to life an aspirational narrative about the future of Britain that is rooted in the hopes and ambitions of people from across our country.

Demos is an independent, educational charity, registered in England and Wales. (Charity Registration no. 1042046)

Find out more at **www.demos.co.uk**

DEMOS

PUBLISHED BY DEMOS MAY 2024

15 WHITEHALL, LONDON, SW1A 2DD
T: 020 3878 3955
HELLO@DEMOS.CO.UK
WWW.DEMOS.CO.UK