Mapping global approaches to public compute

Understanding the options available to policymakers

Compute, or computational power, has emerged as a fundamental currency of AI technologies, shaping not only the technical capabilities of AI systems and who is able to build them, but also the competitive dynamics and strategic landscape of the AI sector. As AI models grow increasingly sophisticated, access to and control of large-scale compute resources have become critical determinants of both research progress and commercial success.

This policy briefing summarises early findings of a research project carried out by the Ada Lovelace Institute (Ada), with the support of the Mozilla Foundation. The research maps existing and planned strategies for public provision of compute with the aim of better understanding the design options available to those scoping and implementing public compute policies.

Ada Lovelace Institute

For more information about the Ada Lovelace Institute and this work, contact please contact the project lead Matt Davies: mdavies@adalovelaceinstitute.org

Introduction

Defining and contextualising public compute

State provision of computing resource is a well-established area of policy, with uses including weather forecasting and climate research. Many of the first 'supercomputers' were procured by public organisations, and today supercomputers operated by the US government and the EU are among the fastest in the world.

The emergence of AI as an industrial priority for countries and trading blocs across the world has raised the prominence of the issue of compute access. New AI-focused initiatives such as the UK's AI Research Resource and the USA's National AI Research Resource pilot aim to provide world-class compute to domestic researchers. Long-established programmes such as the EU's EuroHPC-Joint Undertaking are establishing new AI-related priorities. This activity extends beyond the UK, USA and EU, with countries like China and India also making significant investments in computing resources.

We use the term 'public compute' to refer to this broad family of policies, which use public funds to provide particular groups with access to compute. Initiatives may also be 'public' in other ways – for example, they may provide open, public access to compute or promote particular public interest research areas.

Proponents maintain that public compute investments are necessary for countries to remain competitive in the context of emerging 'arms race' dynamics between states looking to develop advanced AI. Some suggest that these could help to broaden access to the resources needed to train and deploy AI systems, bridging the 'compute divide' between the largest tech companies and smaller companies or research institutions. With increasing calls for 'public' alternatives to corporate AI infrastructure at all levels of the stack, compute – as a central input for modern AI development – is necessarily a part of this picture.

Critics have raised concerns about the scale of public compute investment in the context of a highly-concentrated sector. Ada explored some of these issues earlier this year in a blog post [1] with Common Wealth,[2] highlighting the need for public compute investments to be made within a broader framework of 'market shaping' policies if they are to yield genuine public benefit.

1　Shearer E, Davies M and Lawrence M, 'The Role of Public Compute' (Ada Lovelace Institute, 24 April 2024) https://www.adalovelaceinstitute.org/blog/the-role-of-public-compute/ accessed 31 October 2024

2　Common Wealth' https://www.common-wealth.org/ accessed 31 October 2024

Mapping public compute initiatives

Building on this work, and with support from the Mozilla Foundation, Ada is now working to map existing and planned strategies for public compute to better understand the options available to policymakers.

Our initial mapping has found a landscape characterised by pluralism and experimentalism, with global policymakers not yet converging towards a single approach or set of best practices for public compute. The aims of public compute strategies are diverse, ranging from economic competitiveness to support for public interest innovation, and can sometimes be in tension with each other. Approaches tend to be shaped by existing industrial strengths and weaknesses, national policymaking traditions and international factors such as export controls.

The table below summarises some of the key features of public compute strategies in jurisdictions across the world.

Jurisdiction	Objectives	Key initiatives	Conditions of access	Type of provision	Target users
China	Build domestic capabilities across compute stack in response to US export controls	Direct provision through state-run regional data centres; City-level compute voucher schemes (40-50% subsidies); Equipment exchange platforms.	Location-based requirements (provincial/ municipal); Varies by local implementation; Registration requirements for large language models	Full compute stack through state-run data centres; Focus on domestic hardware/software integration	AI startups and large tech companies; Regional and national research companies
European Union	Build a 'world-class supercomputing ecosystem'; Foster AI development in strategic sectors; Develop European hardware capabilities	Direct procurement and operation of supercomputers through Euro-HPC Joint Undertaking	Must be EU/ participating state organization; Research projects get free access; Publication requirements for research	Supercomputers; Hardware development through the European Processor Initiative; Networking infrastructure	Academic researchers; Public sector organisations; Private companies (esp. SMEs) in strategic sectors
France	Digital sovereignty; Scientific research support; Industrial innovation	Direct procurement and operation of supercomputers through Grand équipement national de calcul intensif (GENCI) investment	Focus on national interests; Strategic sector alignment; Public-private partnership requirements	Supercomputers; Specialised AI compute; Quantum computing infrastructure	Research institutions; Private companies in sectors of national strategic importance; Public agencies
India	Market development; Domestic industry support; Distributed compute access	$600m subsidies for GPU procurement; Open Compute Cloud initiative	Prevention of resale; Government oversight; Market-based allocation	GPUs; Micro data centres; Distributed compute infrastructure	Private companies; Startups; Universities; Industrial users

Jurisdiction	Objectives	Key initiatives	Conditions of access	Type of provision	Target users
United Kingdom	Scientific research advancement; AI innovation support; Research infrastructure development	Direct investment in AI Research Resource (AIRR); Public-private partnerships; Research council funding	Research merit-based access; Academic focus; Project-specific requirements	Supercomputers (e.g., ARCHER2); Cloud computing resources; Research infrastructure	Academic researchers; Scientific community; Selected industry partners
United States	Maintain global AI leadership; Support research innovation; Hardware advancement	National AI Research Resource (NAIRR) pilot delivered through public-private partnerships; DoE Frontiers in Artificial Intelligence (FASST) program	Merit-based research access (NAIRR); Agency-specific requirements	Supercomputers; AI-specific compute; Research infrastructure	Academic researchers; National labs; Government agencies; Industry collaborators

Challenges for public compute initiatives

While these strategies are diverse, we can identify common challenges for public compute initiatives:

Avoiding 'value capture': Value capture is the risk of public investment primarily benefiting private interests, either through direct use of facilities or through the commercialisation of research outputs. The high cost of compute investments creates questions about how to balance equitable and open access for research with returns to the public purse. While public compute initiatives aim to serve researchers and advance innovation, the concentrated nature of the AI sector creates risks that their benefits will primarily accrue to large private companies. Though mechanisms like open publication requirements could help ensure broader public benefit, these remain underdeveloped.

Achieving strategic coherence: In many jurisdictions, multiple overlapping public compute initiatives exist, contributing to a crowded institutional landscape. This presents coordination challenges, particularly around access to crucial complementary resources like data and technical expertise. Without addressing these wider ecosystem constraints through coordinated policy interventions, simply providing hardware access is unlikely to achieve broader innovation and competitiveness goals.

Balancing flexibility and longevity: Public compute strategies are often flexible in response to uncertainty surrounding AI's future capabilities and market structure, making long-term planning difficult. Policymakers must balance the need to remain responsive to rapid technological change against the stability required for long-term infrastructure investment, with uncertainty threatening to undermine confidence and deter private sector engagement.

Squaring compute investments with environmental goals: Public compute infrastructure has significant environmental costs through high energy and water consumption for data centres, potentially conflicting with climate and environmental goals. For example, Ireland's data centre energy usage could account for 28% of national electricity demand by 2031, threatening the country's carbon reduction targets.[3] Policymakers are actively considering strategies to mitigate this, but it is likely that more robust measures will be needed to prevent data centre buildouts from compromising climate objectives.

Methodology and evidence base

Where not otherwise cited, claims made in this briefing are derived from three sources:

• Desk research carried out by the authors.

• Interviews with more than 20 policymakers and experts working on the design, delivery and scrutiny of public compute jurisdictions in jurisdictions including China, the EU, India, the UK and the USA.

• Two expert roundtables with civil society and academic experts.

Our final report, for publication in the next few months, will go into more detail about our research questions and list the individuals we spoke to.

Defining 'public compute'

Compute, or computational power, is used for many applications, but has become most prominent in recent years as a crucial part of the AI supply chain. It is often referred to as a discrete layer within this supply chain, but can refer to a number of different things.

Aligning with the description in *Computational Power and AI* (West and Vipra), we define compute systems as comprising a stack of hardware, software and infrastructure components:[4]

• **Hardware:** chips such as graphics processing units (GPUs), originally used to render graphics in video games but now increasingly applied in AI due to their ability to support complex mathematical computations.

---

3   'EirGrid and SONI (System Operator for Northern Ireland), 'Ireland Capacity Outlook 2022-2031' (2022) https://cms.eirgrid.ie/sites/default/files/publications/EirGrid_SONI_Ireland_Capacity_Outlook_2022-2031.pdf accessed 31 October 2024

4   Vipra J, 'Computational Power and AI' (*AI Now Institute*, 27 September 2023) https://ainowinstitute.org/publication/policy/compute-and-ai accessed 31 October 2024

- **Software:** to manage data and enable the use of chips, and specialised programming languages for optimising chip usage.

- **Infrastructure:** other physical components of data centres such as cabling, servers and cooling equipment.

We use the term 'public compute' to refer to a broad family of policies, which use public funds to provide particular groups with access to compute. These initiatives may also be 'public' in other ways. For example, the Mozilla Foundation has articulated three criteria for 'public AI':[5]

- **Public goods:** Public AI should create open, accessible public goods and shared resources at all levels of the AI technology stack, especially where cost, scarcity and private lock-up pose the greatest barrier to public participation.

- **Public orientation**: Public AI should centre the needs of people and communities, especially those most underserved by market-led development, such as workers, consumers and marginalised groups.

- **Public use:** Public AI should prioritise AI applications in the public interest, especially those neglected by commercial incentives and those posing ethics or security concerns that make them irresponsible to pursue via private development.

As the sections below indicate, some of these features can be found in existing public compute initiatives – but not always, and not consistently.

Public compute initiatives today

State provision of computing resource is a well-established area of public policy, with uses including weather forecasting and climate research. Many of the first 'supercomputers' were procured or built by public organisations, and today supercomputers operated by entities associated with the US government and the EU are among the fastest in the world. Growing interest in AI as an industrial priority has led to the establishment of new AI-focused initiatives, and new priorities for these legacy institutions.

These newer public compute initiatives are characterised by pluralism in aims and policy design: we are not yet seeing clear convergence towards a single approach or set of best practices.

This reflects the fact that public compute policies are not 'one size fits all'. Different delivery models are suited to different aims, from addressing market concentration (France and China), to improving

---

5 'Public AI - Mozilla Foundation' https://foundation.mozilla.org/en/research/library/public-ai/ accessed 31 October 2024

access for research (USA), promoting innovation and business growth (Singapore), and diversifying economies (Saudi Arabia).

The remainder of this section lays out a provisional framework for understanding the different types of public compute initiatives that currently exist.

Design choices

Aims

Public compute initiatives pursue various objectives, often simultaneously. These commonly include the following:

- Supporting scientific research and innovation in general, or in specific research domains (for example, climate science, drug discovery).

- Increasing access to AI research capabilities beyond the private sector.

- Building strategic technological capabilities (for example, in the semiconductor supply chain) through strategic procurement.

- Supporting domestic industry and startups carrying out fundamental AI research.

Traditional research computing infrastructure, exemplified by facilities like ARCHER2 in the UK, primarily focuses on supporting academic research and large-scale scientific modelling.

However, interviewees emphasised that newer initiatives often have broader ambitions around AI development and innovation. The US National AI Research Resource pilot, for instance, explicitly aims to 'democratise' access to AI by providing users with the compute, skills and other resources they need to carry out AI research. India's initiatives emphasise building technological capabilities in strategic sectors and supporting the development of domestic compute supply chains.

Users

Different public compute approaches use different eligibility criteria to determine who can use that resource. This may include:

- public, private or non-profit status

- organisation size

- national or regional status

- academic or scientific merit of user institution

- potential for research to contribute to innovation in strategic areas.

In practice, academic researchers remain the primary users of many public compute resources, though there is a trend toward supporting wider access. The EU's EuroHPC initiative, for example, reserves 50% of resources for EU-based researchers while giving member states flexibility to control allocation of the remaining capacity. India's initiatives target both research institutions and private sector entities, though interviewees noted challenges in defining appropriate access criteria and ensuring equitable distribution of resources.

Hardware

Jurisdictions vary significantly in how much of the 'compute stack' they provide. Models include the following:

- **Full-stack provision** – comprehensive computing services including all elements above

- **Hardware-only support** – direct provision of chips or computing resources

- **Hardware and software combinations** – providing compute resources with necessary software tools

- **Enabling infrastructure provision** – focusing on land, power and other supporting elements

Media coverage of public compute initiatives often focuses on massive-scale compute installations. Traditional high-performance computing (HPC) users in academia often need large-scale systems optimised for modelling and simulation work, which has different architectural requirements than workloads for training or running AI models.

However, our interviews highlighted that users' actual needs are often more modest than this. Most researchers and startups primarily need access to small clusters of GPUs for development work rather than massive parallel computing capability, but might require better support infrastructure, software environments and skills development.

AI-focused companies typically prefer systems optimised specifically for AI workloads, with architectures built around GPUs and modern software stacks. Experts we spoke to suggested that these projects often need flexible, on-demand access rather than having to bid for time on large shared systems. For smaller companies and startups, access to even modest GPU resources can be transformative, especially given the cost of accessing cloud resources at market rates.

Interviewees indicated that government agencies and public sector organisations have particular needs which are different to those of other user groups. They often need compute for specific applied use cases rather than pure research, which often creates requirements around security and data governance.

Provisioning models
The mechanism through which public compute is delivered varies considerably across jurisdictions. Key approaches include the following:

- **Direct public provision**: the state procures, owns and operates compute infrastructure directly

- **Arms-length provision**: delivery through independent entities that may be publicly funded but operate autonomously

- **Public–private partnerships**: collaboration with private firms to provide compute services

- **Private provision with public funding**: private entities provide resources with public subsidy

- **Direct subsidies or vouchers**: financial support for organisations to access private compute

Most existing initiatives follow hybrid models that leverage private sector expertise while maintaining public oversight. This reflects practical considerations around cost and operational expertise, but risks relinquishing public control and jeopardising alignment with policy objectives.

Usage conditions
Access policies often reflect broader strategic priorities around innovation, security and economic development. Common conditions include:

- development practices and safety requirements

- licensing and commercial model restrictions

- requirements around open publication of research

- commitments to public benefit applications.

Our interviews indicated that in most jurisdictions conditionality is being actively explored, with open publication and the adoption of particular licences the most common requirement. At present, profit-sharing or equity arrangements and restrictions on downstream commercialisation are relatively underutilised, reflecting the tendency of these initiatives to focus on basic research rather than lucrative commercial applications.

Costs

Costs vary enormously depending on scope and delivery model, and can typically be funded either through direct public investment, through public-private agreements or through other approaches such as user charging. Traditional supercomputing facilities can require investments in the hundreds of millions or billions, while targeted subsidy programmes might operate at much smaller scales. The UK's cancelled investment in an exascale supercomputer, estimated at around £900 million, illustrates both the scale of investment required and the political challenges of securing long-term funding commitments.[6]

Case studies

Public compute policies are heavily influenced and circumscribed by exogenous factors, including the international environment, domestic market dynamics and the relative strengths of each country. Countries often tailor their public compute strategies based on their existing policy frameworks and areas of comparative advantage. This means that national public compute policies reflect broader geopolitical, economic, and technological conditions.

The following case studies show how different jurisdictions are approaching the provision of public compute, and indicate some of the relevant 'shaping factors'. For each country or bloc we have listed key features of public provision, and factors shaping the continued development of strategy.

China

China's approach to public compute is characterised by significant decentralisation and multiple overlapping initiatives at different levels of government:

- Implementation of state and municipal compute initiatives, including direct provision of compute resources through state-run data centres and regional AI labs.

- City-level compute voucher schemes offering 40–50% cost reductions for AI startups and research.

- Novel coordination mechanisms like government-supported equipment exchange platforms to address supply chain bottlenecks.

- Strategic 'Eastern Data, Western Computing' initiative to build compute centres in western provinces leveraging cheaper energy costs.

- State-level subsidies increasingly focused on equipment and materials rather than just chip design/fabrication.

---

6   'Britain's Government Pulls the Plug on a Superfast Computer' The Economist https://www.economist.com/britain/2024/08/22/
britains-government-pulls-the-plug-on-a-superfast-computer accessed 31 October 2024

China's approach is shaped by the following:

- International context: responding to export controls by building domestic capabilities across the full computing stack while maintaining access to global technology where possible.

- Domestic priorities: balancing support for large tech companies who can stockpile chips with smaller players who face greater constraints under export controls.

- Policymaking approach: significant provincial/municipal autonomy in implementation within broader national strategic frameworks, enabling local experimentation while maintaining central coordination.

- Technical challenges: growing focus on developing software stacks for domestic hardware and building skills for networking compute systems.

European Union
The European High Performance Computing Joint Undertaking (EuroHPC JU), established in 2018, is a collaborative initiative involving the EU, 32 European countries, and three private partners. Funded 50% by participating states and 50% by the EU budget, EuroHPC JU aims to develop and maintain a world-class supercomputing ecosystem in Europe.

Key features relevant to AI include the following:

- Acquiring, upgrading and operating AI-dedicated supercomputers.

- Providing open access to these resources for both public and private entities, with free access for research projects but limited to organisations in EU or participating states.

- New focus on AI model training in strategic sectors like healthcare, climate change, robotics and autonomous driving, redeploying existing allocated resources to support this expansion.

- Emphasising support for startups and medium-sized enterprises, with special access conditions for these entities.

- Developing European hardware capabilities through initiatives like the European Processor Initiative.

The EU approach is shaped by several key factors:

- International context: positioning Europe as a competitive force in global AI development.

- Policymaking comfort zone: leveraging existing multi-national collaborative frameworks (for example, Horizon Europe).

- Adaptation challenges: balancing traditional HPC users with new AI-focused users, while prioritising strategic sectors and smaller enterprises.

France
- France's approach to public compute is characterised by significant national investment, strategic partnerships and a focus on digital sovereignty:

- Major investments in high-performance computing (HPC) infrastructure, including GENCI, to support scientific research and industrial innovation.

- Implementation of a National AI Strategy, launched in 2018, to increase access to compute resources for AI development and research, and the establishment of AI-focused institutes like INRIA to support compute-heavy AI and machine learning projects.

- Continued participation in European initiatives like EuroHPC to build and operate world-class supercomputing infrastructure.

- Emphasis on digital sovereignty through projects like Gaia-X, ensuring critical digital infrastructure remains under national or European control.

- Direct funding and incentives through initiatives like France Relance and Plan Quantum to boost investments in computing infrastructure, AI and quantum computing.

France's approach is shaped by:

- International context: positioning France as a leader in AI and HPC while maintaining digital sovereignty.

- Domestic priorities: aligning compute policies with broader industrial strategy to support key sectors like healthcare, aerospace, and automotive.

- Policymaking approach: balancing national investments with public-private partnerships and participation in EU-wide initiatives.

India

India is pivoting towards a more flexible, market-focused strategy:

- A planned investment of approximately Rs. 4,568 crores (£417,000) over five years to partially subsidise GPU procurement costs for Indian companies.

- Companies benefiting from the subsidy will have the freedom to choose their preferred chips while availing of the subsidy.

- Government oversight to prevent misuse or resale of subsidised compute resources.

- Implementation of an empanelment (selection) process for compute providers to join the public initiative, with current requirements favouring larger, established players.

- The Open Compute Cloud (OCC) project, a non-governmental initiative, which aims to create an alternative supply pool of compute resources, working with smaller providers.

The overall strategy represents a shift from an earlier plan to build a single 10,000-GPU cluster through a public-private partnership.

India's approach is shaped by several key factors:

- Finding a policymaking comfort zone: preference for subsidising purchases rather than direct provision.

- Supporting domestic industry: India's startup, university and industrial ecosystem is well-developed along some parts of the AI value chain and can benefit from these policies.

- Research gaps: lack of comprehensive data on current compute capacity and future requirements, complicating policy-making.

- Longer-term aims: growing interest in developing micro and edge data centres across India, moving away from centralised facilities, and in green data centre initiatives, with some Indian companies attracting international attention for environmentally friendly designs.

- Complementary policies: India has announced production - and design-linked incentive schemes to encourage domestic design and fabrication of advanced semiconductors. So far, it has found success in attracting assembly, testing and packaging activities and memory chip designers.

United Kingdom

The UK's approach to public compute is characterised by a combination of direct investment, public–private partnerships and strategic initiatives aimed at supporting AI research, innovation and scientific advancement:

- Investment in HPC facilities like ARCHER2 to support advanced scientific research across various sectors.

- Implementation of a National AI Strategy emphasising widespread access to computing resources for AI research and development.

- Exploration of a National AI Research Resource to provide public access to computing infrastructure, software and datasets for AI research.

- Collaboration with private sector companies to enable access to cloud-based compute resources and foster innovation.

- Direct investment and targeted subsidies through bodies like UK Research and Innovation (UKRI) to support specific industries and research areas.

The UK approach is shaped by the following:

- International context: positioning the UK as a leader in AI and scientific research on the global stage.

- Domestic priorities: focusing on key sectors such as climate science, energy, and biosciences.

- Policymaking approach: fiscal constraints meaning relatively small (in global terms) investment and political churn creating lack of long term clarity on approach.

USA

The USA's approach to public compute is characterised by a mix of government-led initiatives, public–private partnerships and regulatory frameworks aimed at advancing access to HPC and AI resources:

- National AI Research Resource proposal which would provide researchers and academic institutions access to AI resources, including compute infrastructure, datasets and software.

- Existing government HPC initiatives through national laboratories, with the Department of Energy (DOE) operating the world's two largest supercomputers (Aurora and Frontier) and proposing new AI-focused investments in the form of the Frontiers in Artificial Intelligence for Science, Security and Technology (FASST) programme.

- Public–private partnerships facilitating access to cloud computing resources and promoting AI and machine learning innovation.

- Diverse funding mechanisms, including direct state funding, public-private partnerships, subsidies and grants to support access to computing resources.

The US approach is shaped by the following:

- International context: desire to maintain AI leadership and technological innovation on the global stage, and improve the position of US companies in the hardware supply chain.

- Domestic priorities: emphasis on pushing the frontiers of hardware, and on supporting research and innovation in key sectors like AI, healthcare and climate science.

- Policymaking approach: balancing government initiatives with private sector collaboration and market-driven solutions.

- Inter-agency dynamics: including tension between different agencies' roles in AI and compute initiatives.

Challenges for public compute initiatives

In spite of the diversity of public compute initiatives, we can identify several common challenges across jurisdictions that require further attention from policymakers.

1. Avoiding value capture and delivering public value

The substantial public investment required for compute infrastructure raises important questions about how to balance broad access with creating returns to the public purse and avoiding capture by a few private beneficiaries. While public compute initiatives primarily aim to serve the research community, their procurement decisions can have significant market-shaping effects on suppliers and users.

Through strategic procurement, these initiatives can potentially advance broader policy goals such as developing domestic supply chains and promoting energy-efficient technologies. However, policymakers need to be realistic about what can be achieved through public compute projects alone. The global nature of semiconductor supply chains, which are dominated by a few key players like NVIDIA, means that for most jurisdictions the goal of 'onshoring' production will be a near-impossibility.

One key challenge is value capture: the risk of public investment primarily benefiting private interests, either through direct use of facilities or through the commercialisation of research outputs. The risk of value capture can be mitigated through careful policy design around access and licensing conditions. Some jurisdictions are exploring requirements for open publication of results or commitments to public benefit applications. However, specific mechanisms remain underdeveloped and many of the appropriate levers will lie with other actors or institutions, such as research funding agencies, universities and regulatory bodies.

2. Achieving strategic coherence

The institutional landscape for public compute is increasingly complex, with initiatives operating at municipal, state and federal levels in many jurisdictions including China, the EU and the USA. This creates significant challenges around coordination and strategic alignment, both between different compute projects and with the broader legal and policy framework for AI development and deployment.

The effective use of public compute resources can be constrained by limited access to complementary factors, particularly high-quality data and specialised technical skills. Interviewees told us that there is a particular shortage of expertise in areas like networking and compute software optimisation. While collaboration with industry can provide short-term access to

these resources, longer-term solutions may require broader policy interventions. These could include targeted skills initiatives, competition policy to address market concentration and public investment in data infrastructure.

The experience of several jurisdictions suggests that realising the full benefits of public compute investments requires careful attention to these complementary factors. Simply providing hardware access, without addressing wider ecosystem constraints, is unlikely to achieve policy objectives around innovation and competitiveness – but the agencies developing these initiatives do not always have the necessary authority or remit to drive these wider goals.

3. Balancing flexibility and longevity

Public compute strategies must navigate significant uncertainty about AI's future development, capabilities and markets. Some jurisdictions have responded by maintaining flexibility in their approach: India, for example, has shifted from direct public provision of compute resources to a subsidy-based model supporting private sector access.

However, this flexibility can come at a cost. Infrastructure investments typically require stability and predictability to attract private sector engagement and investment. The recent partial cancellation of the UK's public compute investment illustrates how policy changes can undermine confidence and deter private sector participation.[7] The backlash from industry and researchers highlighted concerns about the UK's commitment to maintaining internationally competitive research infrastructure.[8]

Finding the right balance between adaptability and stability presents a key challenge for policymakers. While strategies need to be responsive to technological change and emerging needs, frequent pivots or reversals can undermine the long-term effectiveness of public compute initiatives.

4. Squaring compute investments with climate and environmental goals

A final challenge relates to the need for these initiatives to square strategic objectives on AI with the significant environmental costs of building compute infrastructure, which may conflict with national climate and environmental goals and, in some cases, legal obligations. Compute infrastructure like data centre facilities have significant energy demands, which is estimated by the IEA to account for around 1–1.5% of global electricity use and 1% of energy-related greenhouse gas emissions.

---

7   Ibid.

8   For example, 'Analysis: The UK Can't Continue Its Shambolic Stop-Go Approach to Supercomputing | UCL News - UCL – University College London' https://www.ucl.ac.uk/news/2024/aug/analysis-uk-cant-continue-its-shambolic-stop-go-approach-supercomputing accessed 31 October 2024

Although global electricity consumption for data centres has grown modestly, countries where data centres are concentrated, like Ireland, have seen data centre electricity usage triple between 2015 and 2023, accounting for 21% of the country's total electricity consumption.[9] EirGrid, Ireland's state-owned electricity transmission operator, predicts that due to significant growth in the data centre market in Ireland, data centre energy usage could account for 28% of national demand by 2031.[10] Ireland's Environmental Protection Agency has stated that it will miss its 2030 carbon reduction targets by a 'significant margin'.[11]

Compute architecture can also consume water at significant rates to assist in evaporative cooling to prevent servers from overheating and in some cases humidity regulation. In Microsoft's latest sustainability report, it acknowledges that its water consumption has increased by 187% between 2020 and 2023 to 7.8 million cubic metres (km^3)[12]. In comparison, Loch Ness in Scotland has a volume of 7.4 km^3.[13]

Investments in public compute risk intensifying these dynamics if they are not coupled with measures to reduce the climate and environmental impact of data centres and their associated infrastructure. Strategies for this are under active consideration by policymakers across jurisdictions. One prominent suggestion is the use of selective procurement to incentivise energy and water efficiency, but it is likely that more robust action will be necessary to ensure that data centre buildouts do not compromise climate and environmental obligations.

Next steps

Over the next phase of this research we will delve into the themes of this briefing in more detail as we work to develop policy options and recommendations for policymakers who are developing public compute strategies.

---

9   'Key Findings Data Centres Metered Electricity Consumption 2023 - Central Statistics Office' https://www.cso.ie/en/releasesandpublications/ep/p-dcmec/datacentresmeteredelectricityconsumption2023/keyfindings/ accessed 31 October 2024

10 EirGrid and SONI (System Operator for Northern Ireland), 'Ireland Capacity Outlook 2022-2031' (2022) https://cms.eirgrid.ie/sites/default/files/publications/EirGrid_SONI_Ireland_Capacity_Outlook_2022-2031.pdf accessed 31 October 2024

11 Agency EP, 'Ireland Projected to Fall Well Short of Climate Targets, Says EPA' https://www.epa.ie/news-releases/news-releases-2023/ireland-projected-to-fall-well-short-of-climate-targets-says-epa.php accessed 31 October 2024

12 '2024 Environmental Sustainability Report, p. 23-34 and data fact sheet, p. 6 | Microsoft CSR' (*Microsoft Sustainability*) https://www.microsoft.com/en-us/corporate-responsibility/sustainability/report accessed 31 October 2024

13 'Freshwater Lochs | NatureScot' https://www.nature.scot/landscapes-and-habitats/habitat-types/lochs-rivers-and-wetlands/freshwater-lochs accessed 31 October 2024

Acknowledgements

The authors are grateful to Andrew Strait, Elliot Jones, Joshua Pike, Michael Birtwistle and Nik Marda for comments on and substantive contributions to this work.