**Host sites for the next wave of UK government AI infrastructure**

**Background**

This funding opportunity is for an expansion of the AI Research Resource (AIRR), a series of public clusters, which will provide compute to AI researchers and startups in the UK. Funding is being provided by DSIT and UK Research and Innovation (UKRI) is managing this process on DSIT’s behalf. This funding opportunity is open to host site providers and their support teams (‘hosting entities’) who have experience and the physical capability of hosting a national large scale compute investment, within the timescales of this funding opportunity.

Up to £500 million of funding is potentially available to support the following costs:

* hardware
* software
* staff
* power
* site development
* other directly relevant costs

We may support more than one site with this funding. For clarity to research organisations this will be paid at 100% full economic cost.

Please note that the total funding available may not all be awarded to the hosting entity directly. Procurement of the necessary supporting technology may be conducted directly through UKRI or HM government procurement processes. No decision on the technology provider has been made in advance of this funding opportunity, which relates solely to the hosting and operation of the service.

Please note that DSIT will retain responsibility for the access policy defining usage of the system. Hosting entities will be expected to work with other AIRR sites.

**Eligibility**

Applications are not limited by organisation type. In order to be eligible for funding, the hosting entity must:

* be able to provide accommodation within a secure UK facility for the operation of the service for a minimum of five years, with the potential to extend for a further year;
* have the space, power and cooling capability to support a large-scale compute system for AI applications that comprises at least 2,000 GPUs or equivalent AI-optimised compute architecture;
* have the ability to deliver infrastructure by March 2025 and operate a full service no later than March 2026; and
* have suitable experience in hosting and operating high quality large-scale compute services for users.

We will not fund:

* sites outside the UK;
* hosts wanting to provide compute provision that is not optimised for AI workloads;
* teams with no experience of hosting and managing large computer infrastructures;
* teams with no knowledge of and experience of working with the wider UK research ecosystem; and
* sites which cannot credibly demonstrate they can deliver to current project timelines.

**Application and assessment process**

Applicants must complete the form below. Completed forms should be sent by email to dri@ukri.org. We must receive your application by **13 February 2024** **at 4:00pm UK time**.

Please note that the information provided here is to enable assessment. Further clarifications and additional data maybe requested by the panel, and by UKRI and/or DSIT, before formal awards are made.

It is intended that your application will be assessed in February 2024 by an independent panel of experts convened specifically for this funding opportunity. Applicants should note that this process may include an interview stage, which will take place w/c 19 February 2024. Further details will be communicated in due course.

Please note that DSIT and/or UKRI reserve the right to reject applications which do not meet the minimum criteria before the panel assessment stage.

Applications will be scored and ranked by the panel, who will make a recommendation to the DSIT Secretary of State, who will make the final decision on funding allocation. Final decisions will be made by DSIT.

**Draft assessment criteria\*:**

1. **Experience in providing large-scale compute services and quality of service**
	1. The host entity’s experience in providing high-quality large-scale compute services, in particular to non-local users.
	2. The capabilities of the host entity, with respect to existing and planned infrastructural provision, security, connectivity, and skills.
	3. Leadership, and suitability of the team.
2. **Viability of the host site**
	1. Whether the host site can deliver the minimum necessary site with appropriate power, space, cooling, and supporting services in the timescale required.
	2. The credibility of the proposed approach to meet the specified requirements, including realistic planning timetables, meeting the specified milestones and financial profiles.
	3. The hosting entity’s proposed approach to managing delivery risks.
	4. Future expansion potential.
3. **Value for money and sustainability**
	1. The degree to which the proposal provides value for money for the duration of the service.
	2. The added value of the hosting entity’s contributions to the operation of the service; in-kind (e.g. existing support infrastructure) or real contributions such as new staff posts).
	3. The quality of the hosting entity’s proposed approach to minimising the environmental impact of the service, including the energy efficiency of the facility.

\* Please note that this is not the final version of the assessment criteria and that final guidance to the review panel may differ from this wording.

1. **Applicant details**

1a. Lead contact details:

|  |  |
| --- | --- |
| Name |  |
| Role |  |
| E-mail  |  |
| Phone |  |

1b. Hosting entity details

|  |  |
| --- | --- |
| Name |  |
| Address |  |
| Legal status |  |
| Is this application being submitted on behalf of a consortium. If so, please include details of the consortium members |  |

1. **The host facility and accommodation**

2a. Summary information:

|  |  |
| --- | --- |
| Maximum number of GPUs or equivalent AI-optimised accelerators that can be accommodated |  |
| If construction is required to meet the eligibility requirements:* when will this be completed
* is planning permission required / applied for / awarded
 |  |
| Power currently available |  |
| Power available by March 2025 |  |
| Power available by January 2026 (until end of service) |  |
| Floorspace available by March 2025 |  |
| Floorspace available by January 2026 (until end of service) |  |
| Location of site |  |

2b. The hosting entity must provide accommodation within a secure UK facility for the operation of the service for a minimum of five years, with the potential to extend. Accommodation must be available to receive hardware by the end of 2024. Please provide details of the accommodation that will be provided, including:

* an overview of the facility;
* details of physical security and fire prevention measures;
* accessibility arrangements for maintenance and incident response, and
* other constraints that may impact on the installation and operation of the service (e.g. floor strength, accessibility for deliveries, etc.).

[Limit of 1000 words for response]

2c. The accommodation must be suitable for the requirements of a state-of-the-art AI compute service comprising at least 2,000 GPUs or equivalent AI-optimised accelerators, including enabling hybrid or direct liquid cooling of the compute hardware, high-speed network connectivity and data storage provision. Please provide details of the physical footprint that will be made available, including details of:

* existing and/or planned hybrid or direct liquid cooling infrastructure, including any modifications and/or limitations of existing infrastructure that would be used in the operation of the service;
* connectivity to internal and external high-speed networks; and
* existing and/or planned data storage provision that would be used in the operation of the service.

[Limit of 1000 words for response]

2d. The hosting entity must make at least 3.0 MW of power available by 1 January 2026 for the operation of the service for a minimum of five years. Please provide details of:

* current and future power provision and power management arrangements that would be made available for the operation of the service. Where upgrades are planned, you should indicate when additional power will be made available. Future upgrades must be based on confirmed upgrade timetables including supporting information from relevant estates teams and power companies; and
* details of what other demands might be made on the total power available within the facility and how this will be prioritised.

[Limit of 1000 words for response]

2e. Future expansion potential of this site will be a consideration of the funding opportunity. Please provide details of future plans beyond 2026, including power, space and staffing. Please articulate realistic figures for the maximal ambition and natural limitations of the site in the next 5-10 years.

[Limit of 1000 words for response]

1. **Experience in providing large-scale compute services and quality of** **service**

3a. The hosting entity will be responsible for operating the service in line with the overall goals of the AI Research Resource to provide large scale accelerator-driven compute according to policies and service standards determined by DSIT. The hosting entity will be expected to provide access to a wide range of key datasets and skilled staff to support this initiative. Please provides details of:

* the hosting entity and leadership team’s track record in hosting large-scale compute services to users, in particular where this relates to AI applications;
* income received by the hosting entity in relation to hosting large-scale compute services;
* technologies currently deployed by the hosting entity, in particular where these relate to accelerator-based large-scale compute and the provision of large-scale deployment of AI;
* the hosting entity’s experience in relation to job scheduling and cluster support software; and
* accreditations (e.g. relevant ISO standards) held by the hosting entity.

[Limit of 1000 words for response]

3b. The hosting entity is expected to provide access to a wide range of key datasets and skilled staff to support this initiative. Please provides details of:

* the hosting entity and leadership team’s track record in supporting and/or conducting cutting-edge AI research, including the skills available within the team;
* current and planned skills profile within the hosting entity, including numbers of full-time equivalents;
* the organisational structure and governance arrangements; and
* plans for recruitment, retention and upskilling of staff.

[Limit of 1000 words for response]

3c. The hosting entity will be responsible for ensuring the security and resilience of the service, meeting agreed service standards, and ensuring a high-quality experience for a wide range of users across the UK. Please provide details of:

* how the hosting entity proposes to maximise the security and resilience of the proposed service and minimise downtime, including cybersecurity and business continuity plans;
* procedures used by the host entity to engage and manage relationships with external users across research base including industry (SME and prime); and
* procedures used by the host entity to support users, including experience of running support services; helpdesk, software support, support for new users with different levels of experience, etc.

[Limit of 1000 words for response]

1. **Sustainability**

4a. The hosting entity is expected to minimise the environmental impact of hosting and operating the service. Please provide details of:

* the energy efficiency of the host facility. Please indicate the average Power Usage Effectiveness (PUE) of the facility, or the planned PUE in the case of new/upgraded facilities;
* the hosting entity’s energy procurement process and policies, including details of any existing long-term contracts that would apply to the operation of the service;
* how the host entity plans to minimise scope one[[1]](#footnote-2) emissions during the operational phase; and
* any other relevant initiatives being undertaken by the host entity.

[Limit of 1000 words for response]

1. **Delivery and Costs**

5a. Please describe plans for delivering the facility to enable commissioning of the full service by March 2026. Please include details of:

* the timeline for delivering the facility, including significant milestone dates relating to construction (e.g. RIBA stages) and/or upgrades;
* any planning permission that is required to deliver the facility (and status of any pending applications);
* any known supply chain constraints, including details of any necessary items with lead times of over six months (excluding compute hardware); and
* significant risks and planned mitigations (maximum of five).

[Limit of 1000 words for response]

5b. Commissioning costs

Please provide non-staff costings for establishing the service below:

|  |  |  |
| --- | --- | --- |
|  | FY 2024/2025 | FY 2025/2026 |
| Site preparation (including. necessary building and/or ground works to prepare the accommodation, including installation of modular data centres | £ | £ |
| Power infrastructure (supply & management, monitoring and backup equipment) | £ | £ |
| Cooling infrastructure (chillers, pipework, etc.) | £ | £ |
| Network provision (within data centre and connectivity to outside networks) | £ | £ |
| Security (physical and cybersecurity infrastructure) | £ | £ |
| Software costs | £ | £ |
| Other (please provide a breakdown and justification in the text box below) | £ | £ |
| **TOTAL** | **£** | **£** |

[Limit of 250 words for response]

5c. Operational and staffing costs

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | FY 24/25 | FY 25/26 | FY 26/27 | FY 27/28 | FY 28/29 | FY 29/30[[2]](#footnote-3) |
| Networking provision | £ | £ | £ | £ | £ | £ |
| Insurances | £ | £ | £ | £ | £ | £ |
| Maintenance (excluding compute hardware) | £ | £ | £ | £ | £ | £ |
| Personnel: infrastructure and system support (inc. cybersecurity) | £ | £ | £ | £ | £ | £ |
| Personnel: software development and maintenance | £ | £ | £ | £ | £ | £ |
| Personnel: user support, training and helpdesk | £ | £ | £ | £ | £ | £ |
| Personnel: security and facility management | £ | £ | £ | £ | £ | £ |
| Personnel: leadership and management team | £ | £ | £ | £ | £ | £ |
| Other (please specify in the text box below) | £ | £ | £ | £ | £ | £ |
| **TOTAL** | **£** | **£** | **£** | **£** | **£** | **£** |

 [Limit of 250 words for response]

5d. Electricity costs

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | FY 24/25 | FY 25/26 | FY 26/27 | FY 27/28 | FY 28/29 | FY 29/30[[3]](#footnote-4) |
| Electricity (per KWh inclusive of taxes). Please indicate current costs and future contractual commitments, where applicable | £ / KWh inclusive of taxes | £ / KWh inclusive of taxes | £ / KWh inclusive of taxes | £ / KWh inclusive of taxes | £ / KWh inclusive of taxes | £ / KWh inclusive of taxes |

5e. Please provide details of the hosting entity’s contributions to the operation of the service; in-kind (e.g. existing support infrastructure) or real contributions such as new staff posts). Justification should be provided for the valuation of contributions.

[Limit of 1000 words for response]

1. Scope one refers to emissions from sources that the hosting entity owns or controls directly. [↑](#footnote-ref-2)
2. Indicative only [↑](#footnote-ref-3)
3. Indicative only [↑](#footnote-ref-4)