# Status of UK AI Supercomputing

Sadaf Alam

Bristol Centre for Supercomputing (BriCS)

University of Bristol

UKLFT Annual Meeting, Edinburgh, April 28-29, 2025

# Bristol Centre for Supercomputing (BriCS) 👶

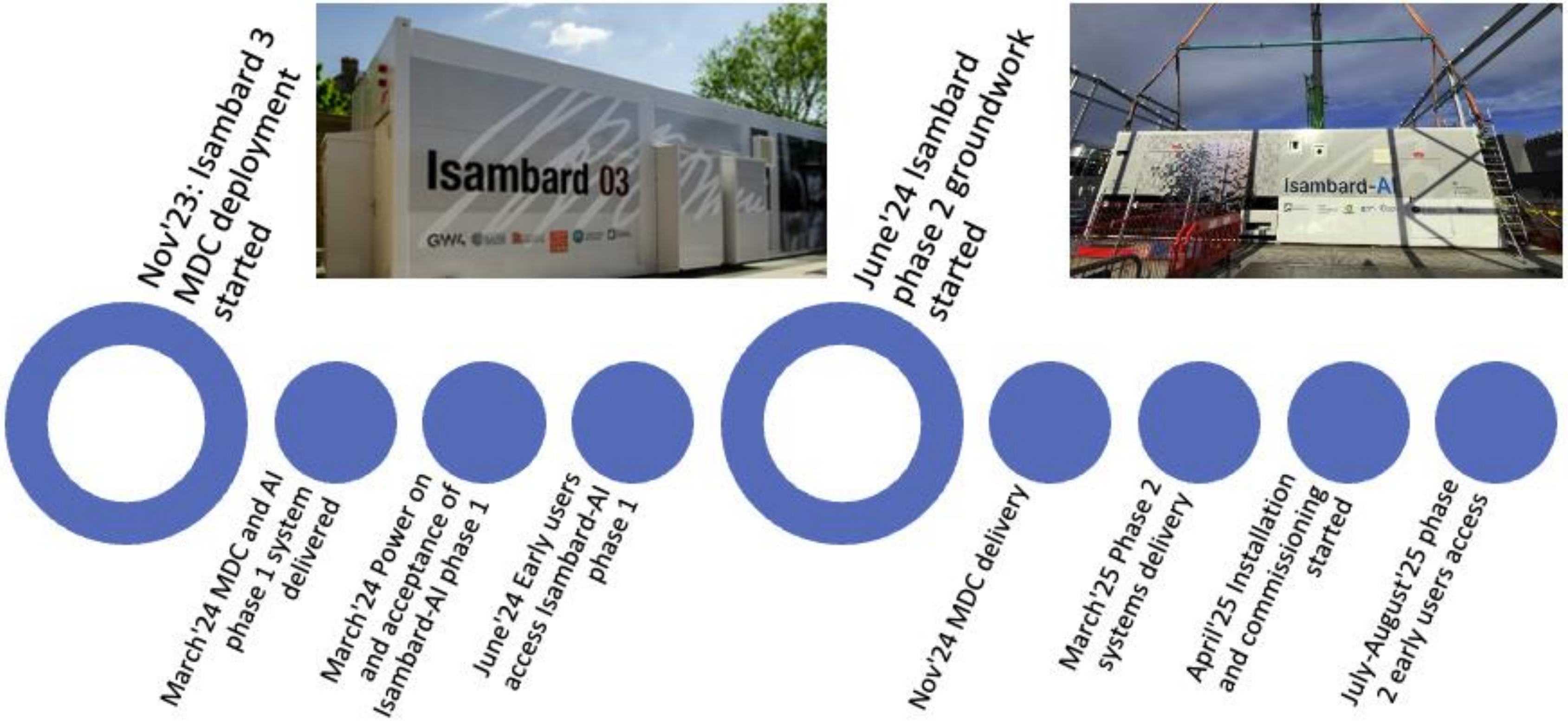
- The GW4 (Bath, Bristol, Cardiff and Exeter) **Isambard project** initially set out to prove that a new **ARM-based** technology was relevant to supercomputing since 2016
- **BriCS** has been recently formed (Q3 2024) for managing Isambard Digital Research Infrastructure (DRI) projects within the Faculty of Engineering
  - Isambard-AI (a GH200 superchip cluster)
    - >£200M including additional Modular Data Centre, all cooling etc.
    - >£300M total investment over 5 years
  - Isambard 3 (a Grace superchip cluster)

BriCS
Bristol Centre for Supercomputing

# Acknowledgements

BriCS (a 🦸‍♂️ 🦸‍♀️ 💥 team)– The work presented in this talk was carried out by various team members 🙏

Isambard-AI is funded by the UK Government's Department of Science, Innovation and Technology (DSIT) via UKRI / STFC. The project has only been possible thanks to a very talented and extremely hardworking team of people across the University of Bristol, HPE, Contour, NVIDIA, the National Composites Centre, and our building contractors, Oakland.



Nov'23: Isambard 3 MDC deployment started

March'24 MDC and AI phase 1 system delivered

March'24 Power on and acceptance of Isambard-AI phase 1

June'24 Early users access Isambard-AI phase 1

June'24 Isambard phase 2 groundwork started

Nov'24 MDC delivery

March'25 Phase 2 systems delivery

April'25 Installation and commissioning started

July-August'25 phase 2 early users access

UKLFT'24

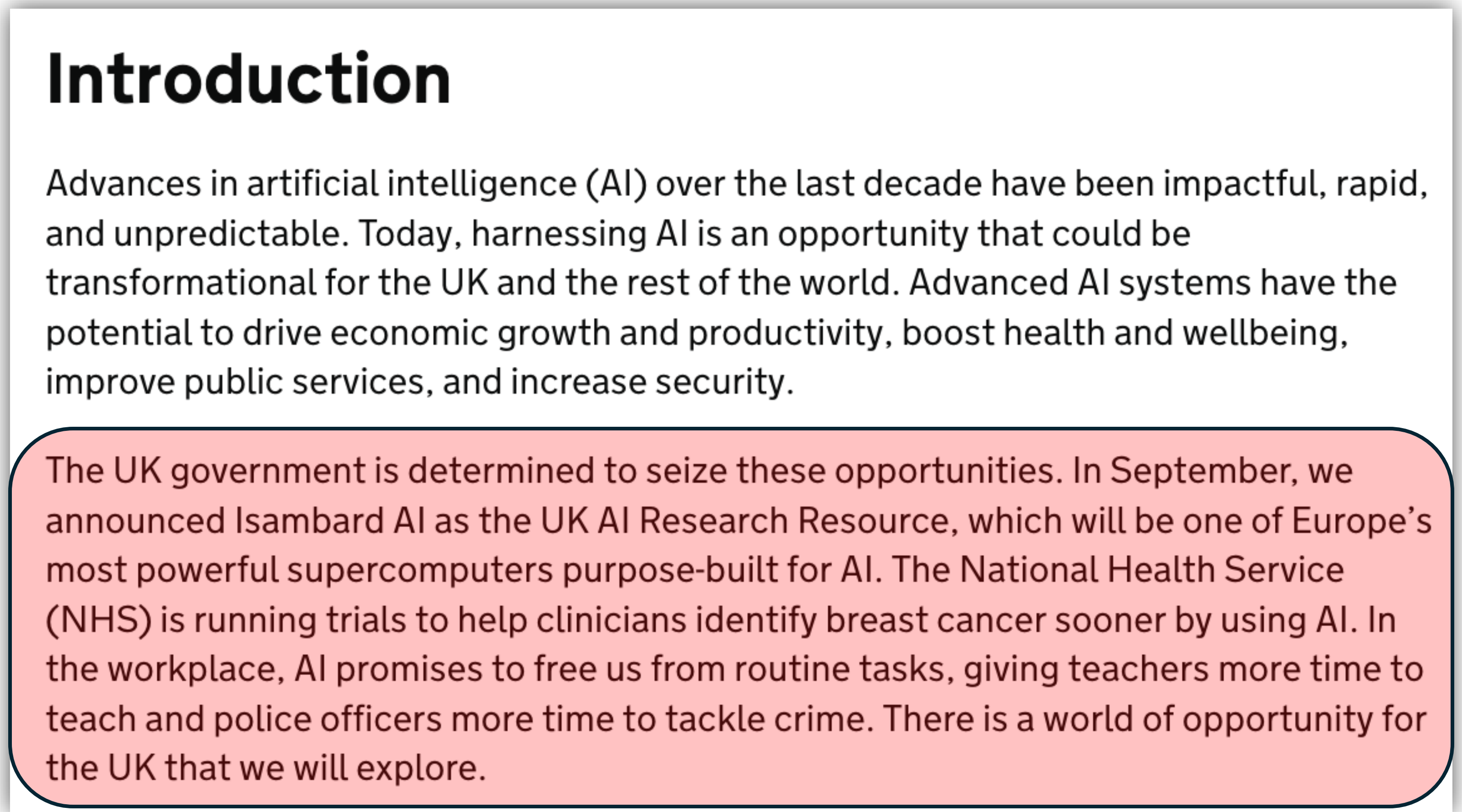
UKLFT'25
😃 my first talk @ UKLFT

# Outline

- UK AI Research Resource (AI RR)
- Isambard-AI (Specifications)
- AIRRFED (Federation)
- AI RR EoI (Access)
- UK AI Action Plan and International Efforts

# UK AI RR = AI Research Resources
# Isambard-AI @ Bristol and Dawn @ Cambridge

Department for Science, Innovation & Technology
AISI
AI SECURITY INSTITUTE
Tackling AI security risks
to unleash growth and deliver the Plan for Change



## Introduction

Advances in artificial intelligence (AI) over the last decade have been impactful, rapid, and unpredictable. Today, harnessing AI is an opportunity that could be transformational for the UK and the rest of the world. Advanced AI systems have the potential to drive economic growth and productivity, boost health and wellbeing, improve public services, and increase security.

The UK government is determined to seize these opportunities. In September, we announced Isambard AI as the UK AI Research Resource, which will be one of Europe's most powerful supercomputers purpose-built for AI. The National Health Service (NHS) is running trials to help clinicians identify breast cancer sooner by using AI. In the workplace, AI promises to free us from routine tasks, giving teachers more time to teach and police officers more time to tackle crime. There is a world of opportunity for the UK that we will explore.

https://www.gov.uk/government/publications/ai-safety-institute-overview/introducing-the-ai-safety-institute (November 2023)

# BriCS outlook (Isambard-AI)

## Phase 1

Arrived & accepted in March 2024 (Isambard 3 MDC)

**1 x DLC EX2500 cabinet**

21 blades (4-way Grace-Hopper)

42 nodes

168 GH superchips

12,096 Neoverse V2 Armv9 CPU cores

168 Hopper GPUs

21.5 TB CPU memory

16.1 TB high bandwidth GPU memory

37.6 TB total memory

**AI high performance storage**

~1 PB all-flash ClusterStor Lustre

## Phase 2

New Isambard-AI ~5MW MDC Delivery of AI services (s)

**12 x DLC EX4000 cabinets**

660 blades (4-way Grace-Hopper)

1,320 nodes

5,280 GH superchips

380,160 Neoverse V2 Armv9 CPU cores

5,280 Hopper GPUs

675 TB CPU memory

506 TB high bandwidth GPU memory

1.18 PB total memory

**AI high performance storage**

~27 PB all-flash storage!

~20 PB Lustre, ~7 PB software defined VAST

The GREEN 500 CERTIFICATE

Isambard-AI phase 1, an HPE Cray EX254n System at the University of Bristol, Bristol, United Kingdom

is ranked

No. 2 in the Green500

among the World's TOP500 Supercomputers

with 68.83 GFlops/Watt Linpack Power-Efficiency

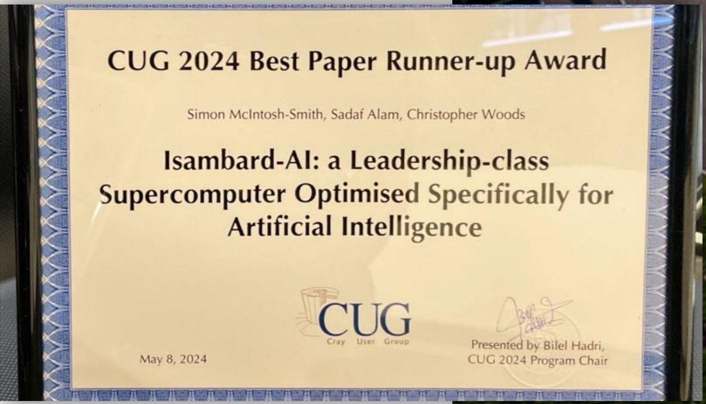
on the Green500 List published at the ISC High Performance, May 13, 2024

Congratulations from the Green500 Editors

Wu-chun Feng
Virginia Tech

Kirk Cameron
Virginia Tech

ISC2025 (Green500)



CUG 2024 Best Paper Runner-up Award

Simon McIntosh-Smith, Sadaf Alam, Christopher Woods

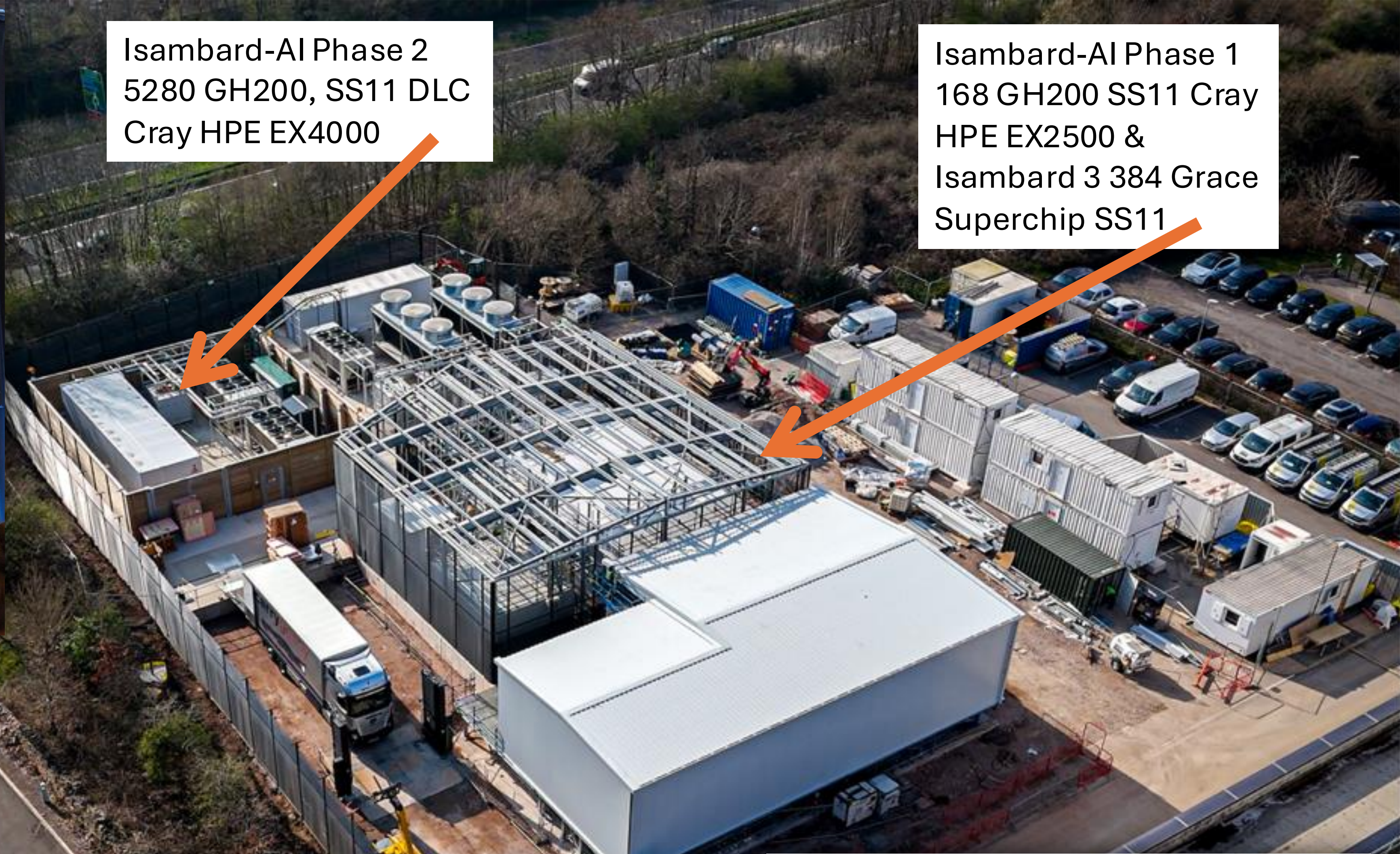
Isambard-AI: a Leadership-class Supercomputer Optimised Specifically for Artificial Intelligence

CUG
Cray User Group

May 8, 2024

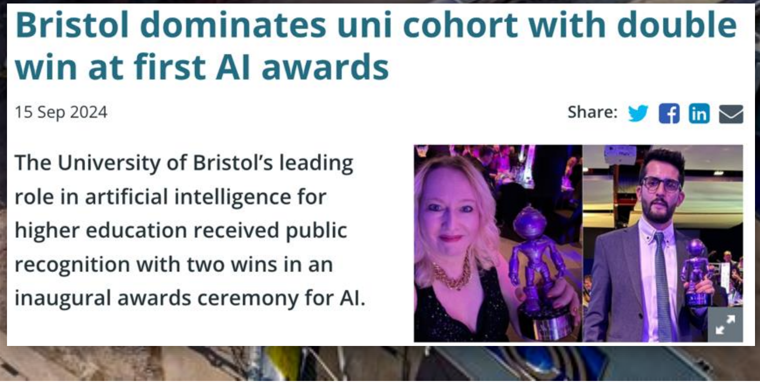
Presented by Bilel Hadri,
CUG 2024 Program Chair

University of
BRISTOL



Isambard-AI Phase 2
5280 GH200, SS11 DLC
Cray HPE EX4000

Isambard-AI Phase 1
168 GH200 SS11 Cray
HPE EX2500 &
Isambard 3 384 Grace
Superchip SS11



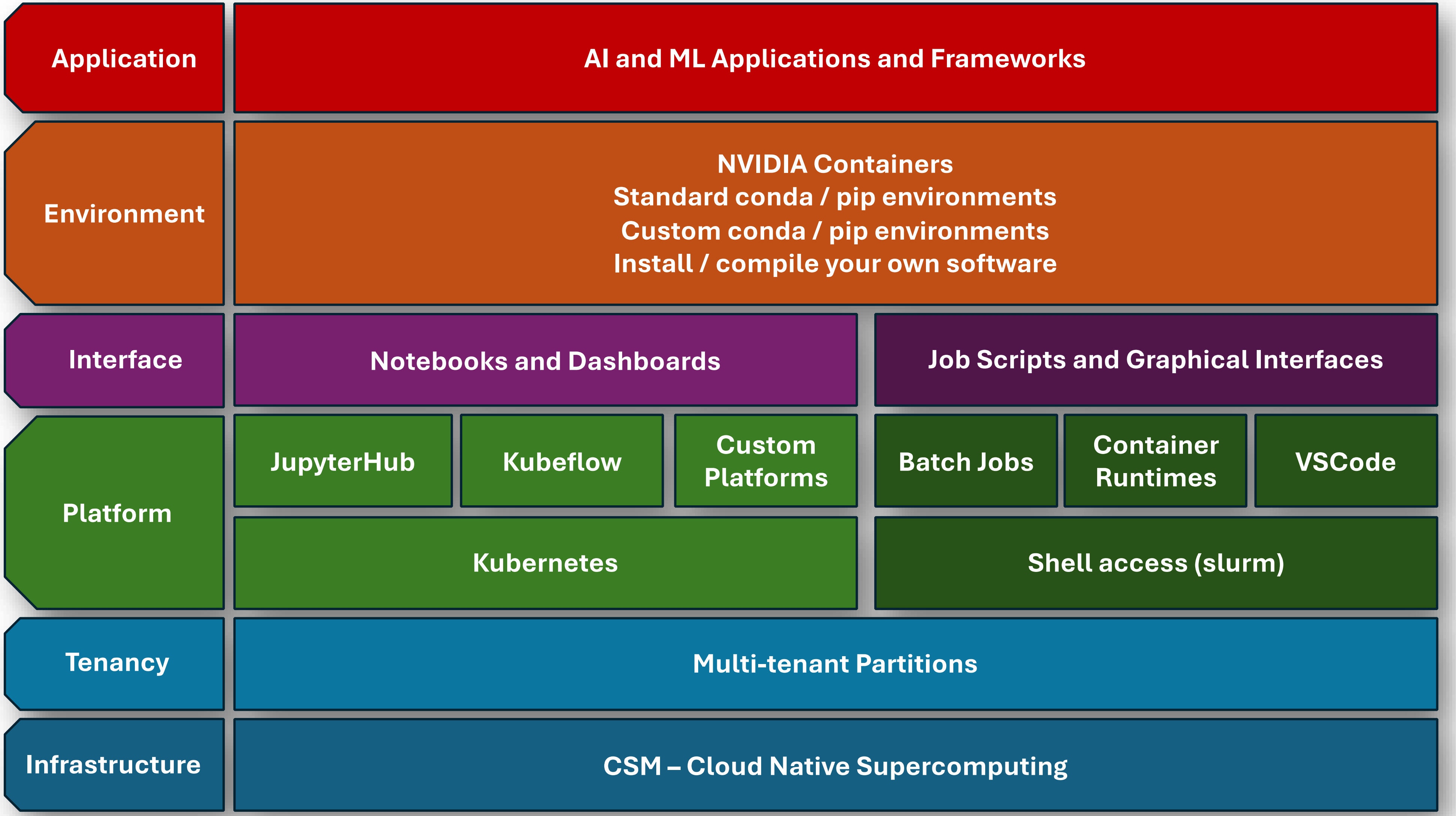
## Bristol dominates uni cohort with double win at first AI awards

15 Sep 2024

Share:

The University of Bristol's leading role in artificial intelligence for higher education received public recognition with two wins in an inaugural awards ceremony for AI.

# Convergence of Cloud and HPC on Isambard-AI



Application: AI and ML Applications and Frameworks

Environment: NVIDIA Containers
Standard conda / pip environments
Custom conda / pip environments
Install / compile your own software

Interface: Notebooks and Dashboards | Job Scripts and Graphical Interfaces

Platform: JupyterHub | Kubeflow | Custom Platforms | Batch Jobs | Container Runtimes | VSCode
Kubernetes | Shell access (slurm)

Tenancy: Multi-tenant Partitions

Infrastructure: CSM – Cloud Native Supercomputing

University of BRISTOL

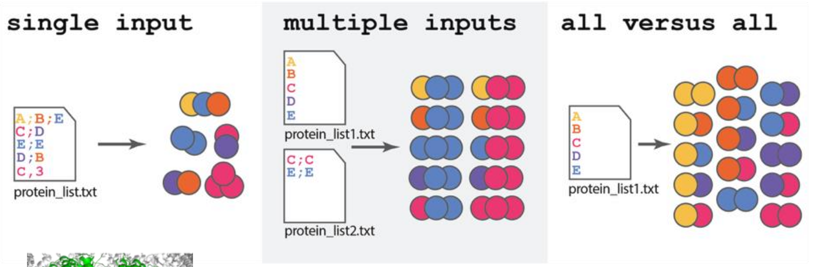
Details in Isambard-AI: a leadership class supercomputer optimised specifically for Artificial Intelligence. https://arxiv.org/abs/2410.11199

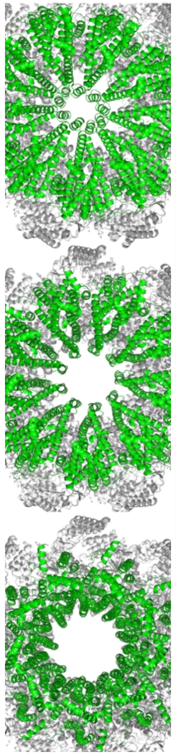
BriCS
Bristol Centre for Supercomputing

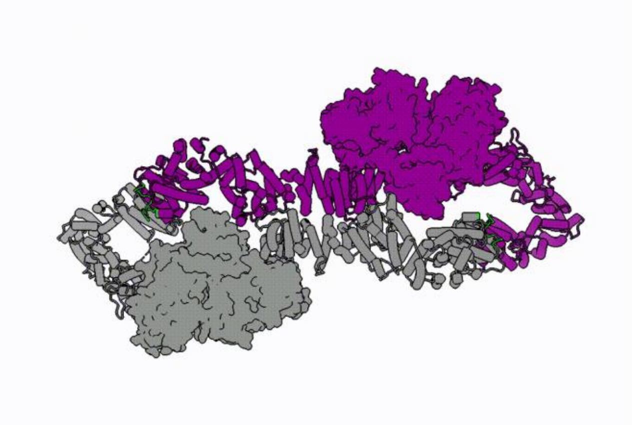
# More Users, Projects & Success Stories

- 100+ projects, ~500 users
- UCL **BritLLM** (UK models specific for law, health, and finance applications, & local languages) https://llm.org.uk, https://arxiv.org/abs/2410.23956
- AlphaFold and OpenFold for **Cardiac disease (**Understanding how gene mutations change protein complexes and cause inflammation)
- Sensitivity of models to backdoor attacks after **data poisoning (**How easy is it for malicious actors to insert poisoned training data for a universal jailbreak?)

  - BIAS-AI (**Bias** of diagnosis systems to skin colour)
  - Turing Institute fully **reproducible model training** including optimiser checkpointing (GPTNeoX)







Nf1: morphing of inpainted and experimental models

# DAWN

256 Dell PowerEdge XE9640 nodes with DLC to CPUs and GPUs

2x 48 cores Intel Xeon 4th gen (Sapphire Rapids) CPU

4x Intel Data Center Max GPU 1550 (Ponte Vecchio) GPUs

4x HDR 200 Gb/s per node

~ 100 TFlops per node

# AIRRFED (AI RR Federation Project)

- A thin federation
  - IdP federation with MyAccessID
- A capacity management portal for the AI RR Marketplace
  - AIRRPortal
- 1-year project 🏎

Goal: AI RR Federation project provides a single-pane of glass control panel or dashboard to AI compute services allocators, researchers, industry users and service providers in a cybersecurity compliance manner

- Status
  - ✓Thin federation completed at both sites including user and project management (maintaining site and service provider autonomy)
  - ✓AIRRPortal demo last month (March 13-14, 2025, workshop), which will be operational in Q3 2025 for allocators

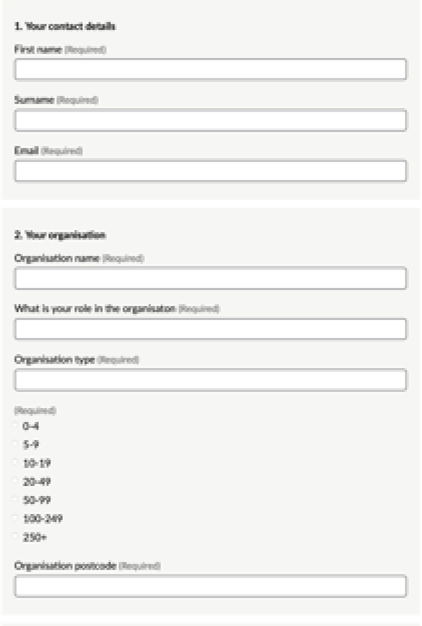
# Motivation: Supercomputing in the Age of AI

- AI research and development advances rapidly
- New, game changing developments appear nearly every week and month
- Poses a challenge for traditional research, which operates on timescales of years
- Particular challenge for managing access to compute
  - Calls can be annual, allocating compute for up to a year at a time
- Research can be scooped and become outdated in the time it takes to apply for and wait for a supercomputing allocation
- How can UK Research And Innovation keep up in the Age of AI?

Credit: Christopher Woods presentation, AIRRFED workshop March 13-14, 2025

## Idea

## Application



**1. Your contact details**

First name (Required)

Surname (Required)

Email (Required)

**2. Your organisation**

Organisation name (Required)

What is your role in the organisation (Required)

Organisation type (Required)

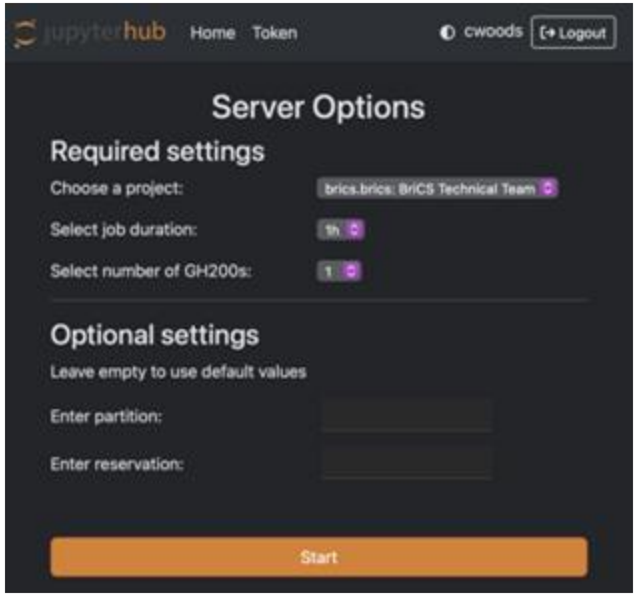
(Required)

○ 0-4
○ 5-9
○ 10-19
○ 20-49
○ 50-99
○ 100-249
○ 250+

Organisation postcode (Required)

## Allocation

## Access



Server Options

Required settings

Choose a project: brics.brics: BriCS Technical Team

Select job duration: 1h

Select number of GH200s: 1

Optional settings

Leave empty to use default values

Enter partition:

Enter reservation:

Days to Weeks – NOT months to years

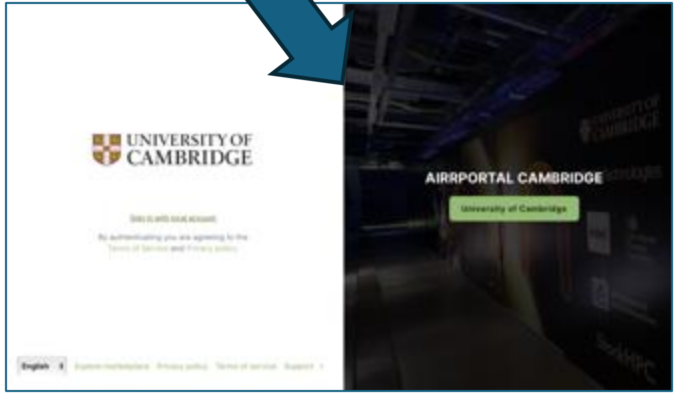
University of BRISTOL

Credit: Christopher Woods presentation, AIRRFED workshop March 13-14, 2025

BriCS
Bristol Centre for Supercomputing

Allocators use AIRRPortal as a single dashboard to allocate projects to the Dawn and Isambard AIRR Systems
Can query / graph / ask LLM to understand how allocations are being used

AIRRPortal communicates the allocator's commands to the Isambard and Dawn sites. Projects are created. Users onboarded. Resource consumed. All reported back to the allocator's AIRRPortal Dashboard



**Single pane of glass empowers allocators to manage and direct AIRR Compute Traffic across national AIRR resources**

University of BRISTOL

Credit: Christopher Woods presentation, AIRRFED workshop March 13-14, 2025

BriCS
Bristol Centre for Supercomputing

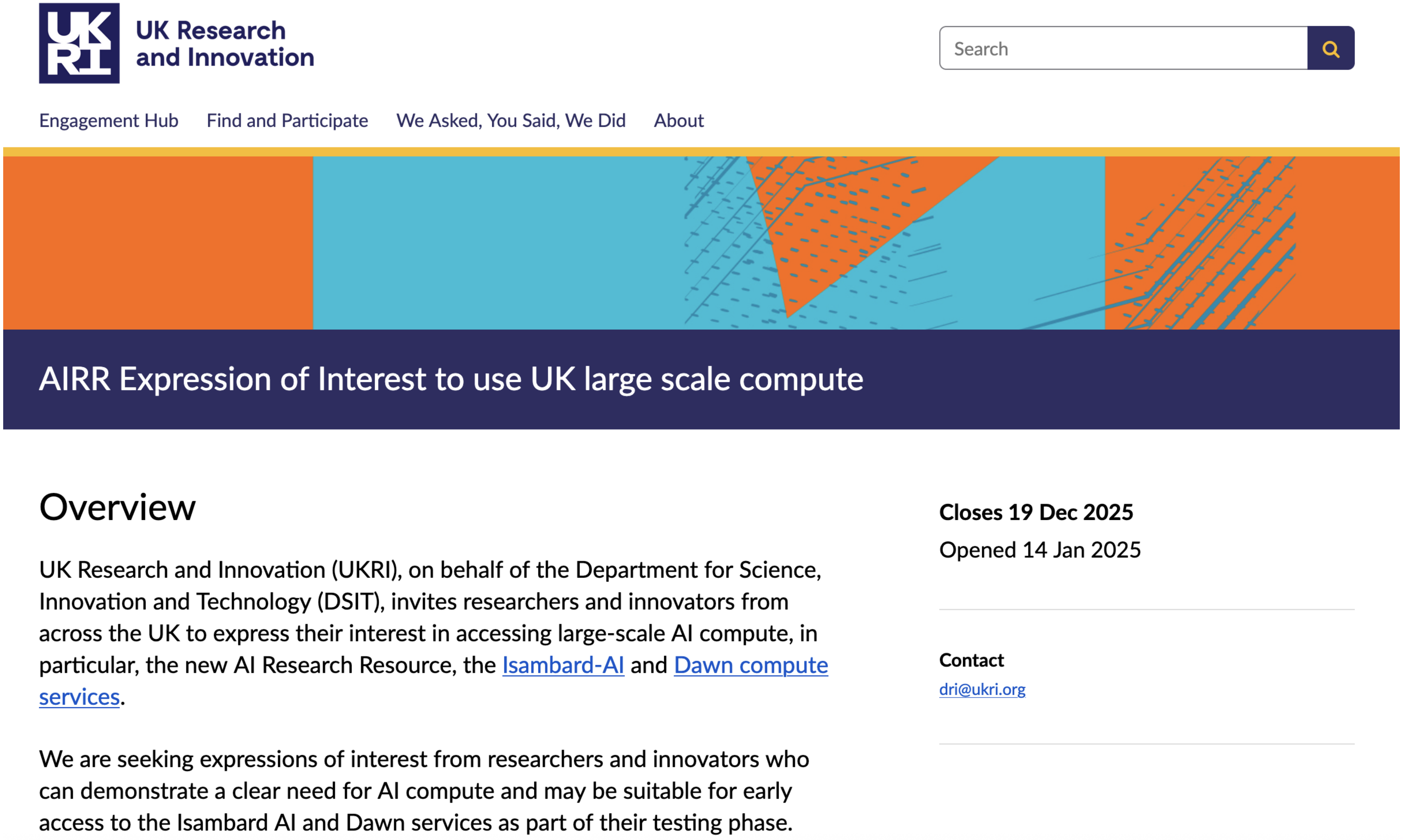
# Opportunity: National Federated Compute Services Flexible Funding Call

| **Total funding for the call** | **£1.98M** |
|---|---|
| **Maximum award** | £200k |
| **Funding level** | 80% of FEC |
| **Call opens** | 25/03/25 |
| **Closing date for final submissions** | 16/05/25 |
| **Notification of outcomes** | 20/06/25 |
| **Anticipated earliest project start date** | 01/07/25 |
| **Project completion deadline** | 31/07/26 |
| **Maximum duration of projects** | 12 months |

https://www.archer2.ac.uk/community/nfcs/

The key benefits that federation can deliver are:

- improved accessibility for users,
- enhanced data accessibility and sharing,
- efficient use of deployed resources through pooling and sharing,
- enhanced sharing of expertise and experience across disciplinary boundaries,
- enhanced cybersecurity through joined-up operations and policy,
- enhanced resilience through the distribution of interoperable resources and services
- larger community of highly skilled researchers and technical staff entering UK workforce



# AIRR Expression of Interest to use UK large scale compute

## Overview

UK Research and Innovation (UKRI), on behalf of the Department for Science, Innovation and Technology (DSIT), invites researchers and innovators from across the UK to express their interest in accessing large-scale AI compute, in particular, the new AI Research Resource, the Isambard-AI and Dawn compute services.

We are seeking expressions of interest from researchers and innovators who can demonstrate a clear need for AI compute and may be suitable for early access to the Isambard AI and Dawn services as part of their testing phase.

**Closes 19 Dec 2025**

Opened 14 Jan 2025

**Contact**

dri@ukri.org

University of BRISTOL

https://engagementhub.ukri.org/ukri-infrastructure/airr-eoi/

BriCS Bristol Centre for Supercomputing

# AI OPPORTUNITIES ACTION PLAN

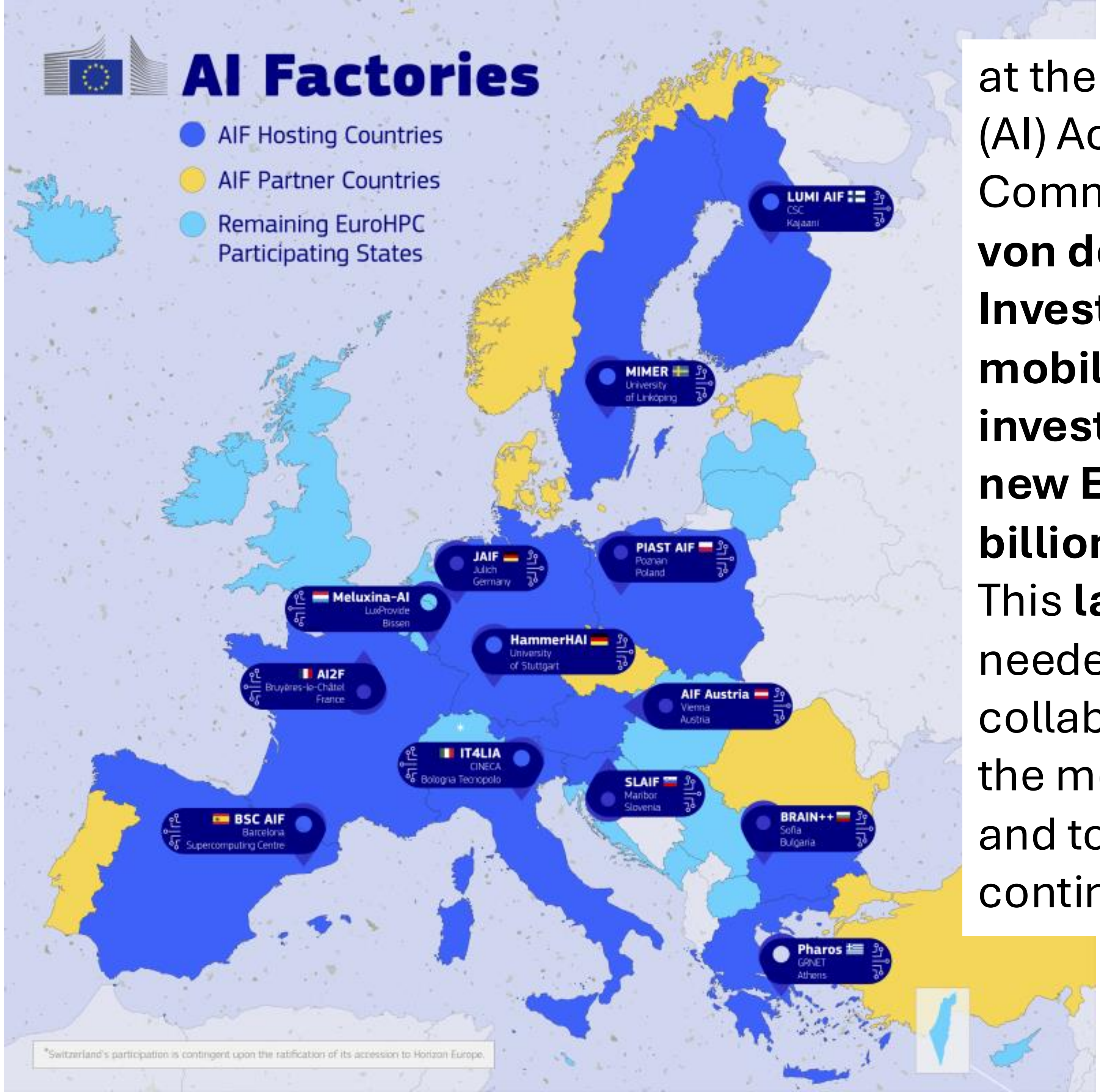
Ramping up AI adoption across the UK to boost economic growth, provide jobs for the future and improve people's everyday lives

AI Opportunities Action Plan. Ramping up AI adoption across the UK to boost economic growth, provide jobs for the future and improve people's everyday lives.

- **Invest in the foundations of AI:** We need world-class computing and data infrastructure, access to talent and regulation (Section 1).
- **Push hard on cross-economy AI adoption:** The public sector should rapidly pilot and scale AI products and services and encourage the private sector to do the same. This will drive better experiences and outcomes for citizens and boost productivity (Section 2).
- **Position the UK to be an AI maker, not an AI taker:** As the technology becomes more powerful, we should be the best state partner to those building frontier AI. The UK should aim to have true national champions at critical layers of the AI stack so that the UK benefits economically from AI advancement and has influence on future AI's values, safety and governance (Section 3).

https://www.gov.uk/government/publications/ai-opportunities-action-plan/ai-opportunities-action-plan

**AI Factories** leverage the supercomputing capacity of the EuroHPC Joint Undertaking to develop trustworthy cutting-edge generative AI models.


AI Factories
● AIF Hosting Countries
● AIF Partner Countries
● Remaining EuroHPC Participating States
LUMI AIF CSC Kajaani
MIMER University of Linköping
PIAST AIF Poznan Poland
JAIF Julich Germany
Meluxina-AI LuxProvide Bissen
HammerHAI University of Stuttgart
AI2F Bruyères-le-Châtel France
AIF Austria Vienna Austria
IT4LIA CINECA Bologna Tecnopolo
SLAIF Maribor Slovenia
BSC AIF Barcelona Supercomputing Centre
BRAIN++ Sofia Bulgaria
Pharos GRNET Athens

*Switzerland's participation is contingent upon the ratification of its accession to Horizon Europe.

at the Artificial Intelligence (AI) Action Summit in Paris, Commission President Ursula **von der Leyen** has launched **InvestAI, an initiative to mobilise €200 billion for investment in AI, including a new European fund of €20 billion for AI gigafactories**. This **large AI infrastructure** is needed to allow open, collaborative development of the most complex AI models and to make Europe an AI continent.

University of BRISTOL

BriCS
Bristol Centre for Supercomputing

# Thank you

- Contact: brics-enquiries@bristol.ac.uk
- https://docs.isambard.ac.uk/