UKRI Science and Technology Facilities Council

AI for Scientific Innovation

STFC AI Strategy

Our vision

Creating and applying responsible AI knowledge and solutions, maximising the value of the data we generate, capture, access and store, for growth and prosperity.

 Cover image credit: Getty Images

Contents

Foreword

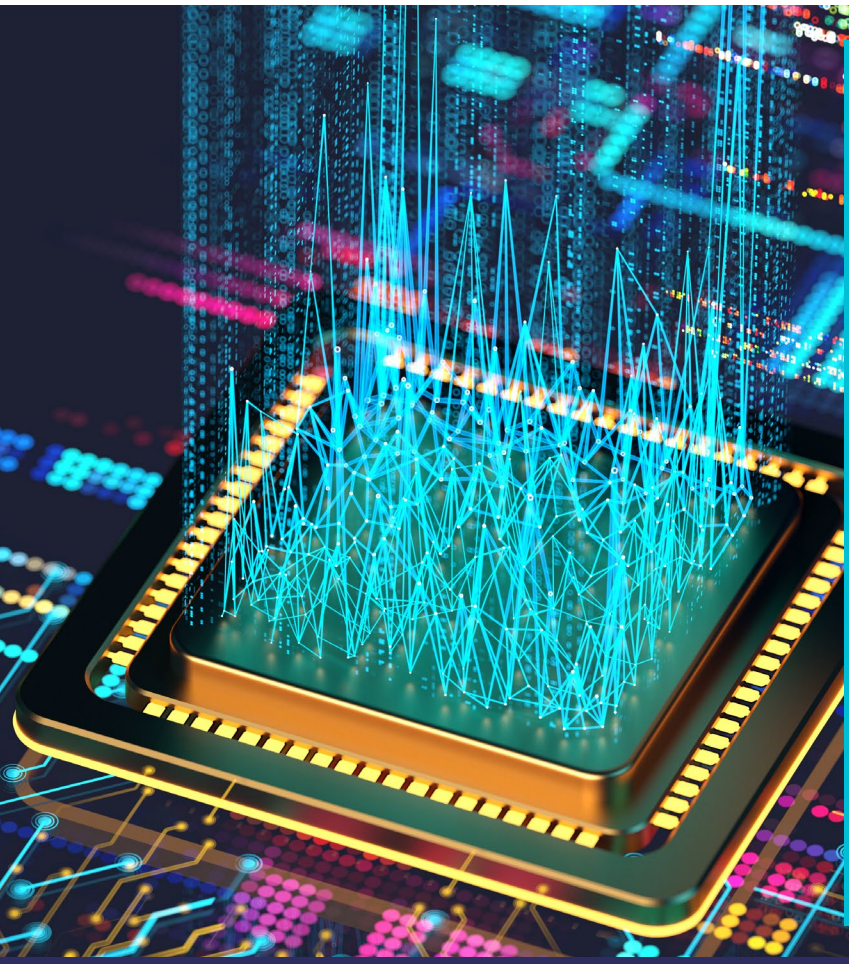
For over three decades, communities supported by the Science and Technology Facilities Council (STFC) have been pioneering the development and application of novel machine learning (ML) and artificial intelligence (AI) solutions to extract maximum value from the huge volumes of data generated across our investments.

During this time, AI has developed into a collection of tools and technologies that are now integral to our core mission of discovering the secrets of the Universe, developing advanced technologies, and innovating to solve real-world challenges. Driving enhanced productivity of science is likely to be "the most economically and socially valuable use for AI"[1].

Strategic advantage in AI will be derived from creating and using AI knowledge and solutions to solve complex research and innovation (R&I) challenges. By translating and exchanging this knowledge and insight across the R&I system into practical applications across the economy, we aim to contribute to the development of sovereign AI capabilities, for growth and prosperity.

We define these collective themes as "AI for Scientific Innovation", and this strategy has been developed to direct the application and adoption and embed this way of working across STFC.

The strategy is in part a response to UKRI's Transforming our World with AI, the Royal Society's Science in the Age of AI and the national AI Opportunities Action Plan. It aims to inspire our science, technology and engineering communities as well as create opportunities to collaborate across UKRI, and with our academic, industry and government partners, policy makers, and funders. We look forward to working together to drive AI for Scientific Innovation.



Predicting proteins with AI

The automated VMXi beamline at Diamond Light Source has made use of AlphaFold, an AI model that predicts a protein structure, as an integral part of our data analysis pipelines. It has facilitated the acceleration of the determination of new protein structures. It allows users to enter in a simple protein sequence when submitting their samples providing a starting structure against which conventional refinement routines can then use when refining against the data collected at Diamond. This not only makes the experience of using the beamline less complex for users but also increases the likelihood of resolving the protein structure. It does all of this whilst reducing computational load allowing for a higher throughput of samples to be examined.

1. oecd.ai/en/wonk/accelerating-science

Image credit: Getty Images

Executive summary

The environment fostered by STFC is characterised by the integration of advanced software, powerful computing capabilities, a legacy of data-intensive scientific research, large and complex data sets, cutting-edge AI technologies and specialised domain expertise. It serves as an incubator for advancing AI-driven scientific innovation as well as innovation in AI, within the UKRI ecosystem.

Advancing STFC's science and technology with AI simultaneously supports UKRI's ambitions to drive innovation in AI, pushing the boundaries of both disciplines.

To maximise our impact as both a consumer and producer of AI technology, we have identified three overarching strategic goals:

1. Using AI to productively advance STFC's science, technology, engineering and innovation, maintaining our status and influence as a world-leading R&I organisation and simultaneously driving AI innovation
2. Securing the health of the AI ecosystem in collaboration across UKRI and the wider R&I ecosystem
3. Enabling organisations across the UK to harness the potential of AI to solve real-world challenges

Acceleration, automation, enhanced data analysis and better, more precise, decision-making through AI all have the potential to transform science. They provide the opportunity to make extraordinary insights accessible, to make the remarkable routine, to deliver science with greater throughput and reduced uncertainty, and to engender approaches that can be translated to make impact in new settings and with new communities. Ultimately, AI creates the potential to move to the next level in our collective science, technology and innovation programme.

A healthy AI ecosystem requires data, compute, people and the development of responsible, trustworthy software. These aspects must operate and be optimised together. The foundations of a healthy AI ecosystem have been laid in the UK with the government's AI Opportunities Action Plan recognising the potential of AI to transform people's lives and the economy. However, the pace of global investment in AI as applied to science is rapidly increasing. This places the UK's and STFC's leading position on the global science stage at risk. For example, in the US the Frontiers of AI for Science, Security and Technology (FASST) programme is part of an investment of $20 billion across 10 years in AI research and development.

STFC is a complex, multidisciplinary organisation dedicated to uncovering the secrets of the Universe. We do this by advancing fundamental physics and advancing and exploiting cutting-edge technologies that underpin fundamental physics and big science. We innovate to address real-world challenges by translating and exploiting our capabilities, know-how and skills across the economy. Within UKRI, we enable and deliver science at scale and, through the UK's National Laboratories and large-scale science facilities, facilitate delivery of the collective ambition of UKRI and the broader R&I ecosystem. STFC's unique capabilities and expertise in compute, skills and data, its robust, agile models of partnership and approach to co-creation and inter-disciplinarity are critical to driving the application and adoption of AI.

There is an interdependence between the physical sciences and AI. Whilst STFC's science challenges are at surface-level domain-specific and complex, they represent an important testbed for a wide variety of new AI techniques.

Building on our core science programme, and within UKRI, we operate the UK's National Laboratories and campuses as an interlinked ecosystem for innovation in science and technology. This accelerates commercialisation and brings competitive advantage and skills for the UK as well as broader industrial, societal and economic impact. In the context of AI, we aim to enable organisations of all types and sizes to exploit UK compute capabilities and explore and adopt digital and AI solutions.

AI for space traffic management

Spaceflux operates a global network of optical sensors that deliver precise space situational awareness (SSA) data, managed through the CORTEX platform. The company was founded in 2022 and currently employs 30 full-time staff. Spaceflux has secured major UK and international government contracts, establishing the company as a leader in space traffic management. Its CORTEX platform uses advanced analytics (including AI-based) that are based on experience developed for the removal of noise in astronomical data and the study of exoplanet atmospheres from the STFC-ESA Ariel mission.

Because AI touches on everything that we do, we have a mission-led obligation to drive AI innovation, in conjunction with colleagues across UKRI, and enable its application and adoption across STFC, the broader UKRI portfolio and wider R&I system. For the full potential of AI to be harnessed, we must take a cross-organisation approach to consolidating, building on and integrating our strengths across the landscape, at scale and pace.

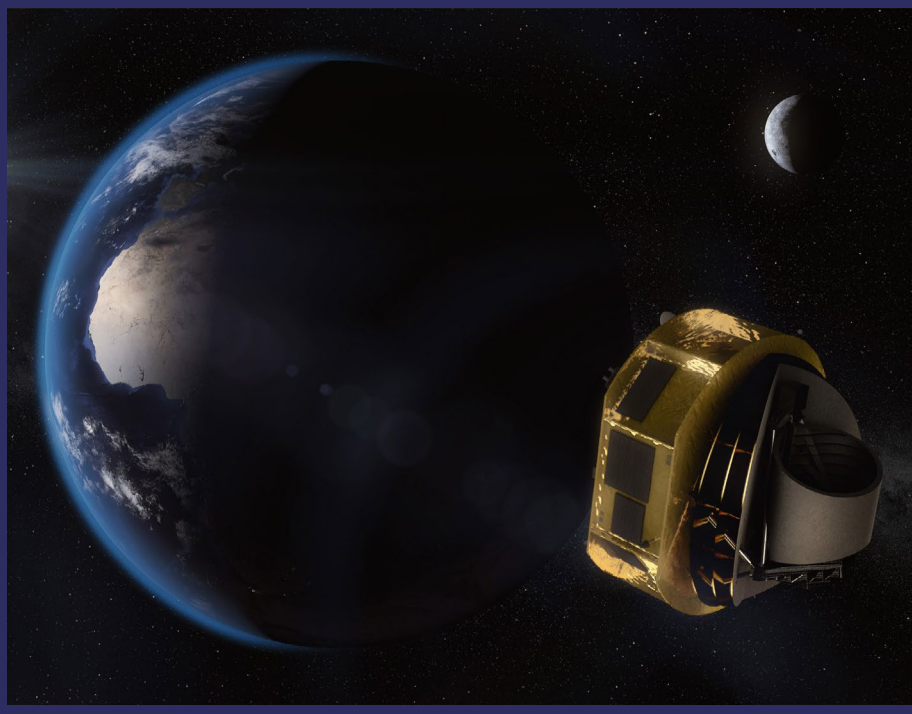
Our strategy is built on five strategic objectives:

- **AI skills**
- **AI-ready data**
- **AI challenge-led technical solutions**
- **Mission-led AI solutions**
- **Responsible AI innovation**

These core objectives are supported by the underpinning objective of effective partnerships.

We have identified four key enablers to support delivery of this strategy:

- **People and governance**
- **Delivery models and mechanisms**
- **Access to compute**
- **Funding**



Artist's impression of Ariel on its way to Lagrange Point 2 (L2).

Image credit: ESA/STFC RAL Space/ UCL/Europlanet-Science Office

Our strategic vision

Our vision is of STFC creating and applying responsible AI knowledge and solutions, maximising the value of the data we generate, capture, access and store, for growth and prosperity.

This AI strategy will enable us to realise this vision. The strategic goals we have identified will help us to accelerate scientific discovery and solve our complex R&I challenges as effectively and efficiently as possible. They will enable us to play a leading role across UKRI in developing and sustaining a healthy AI ecosystem for the benefit of our core investments and communities as well as deploy our capabilities and expertise for the benefit of the wider R&I ecosystem. Critically, they will guide us to translate and exchange our unique knowledge and insight into practical applications across the wider R&I ecosystem and economy.

Our strategic goals

Strategic goal 1: Using AI to productively advance STFC's science, technology, engineering and innovation, maintaining our status and influence as a world-leading R&I organisation and simultaneously driving AI innovation

It is widely acknowledged that AI is an integral, foundational set of tools and technologies required to advance our science and technology. It provides us with new ways to derive value from the complex, diverse data generated and stored across STFC. AI will enable complex challenges to be addressed productively, generate new knowledge from existing data, achieve greater scientific throughput, and open up new lines of enquiry.

In transforming our own core science and technology using AI we will necessarily push the boundaries of AI.

Across STFC we hold approximately 500 petabytes (PB) of data across disciplines – by far the largest coherent dataset held in the public sector. Approximately 75% of this is generated from scientific instruments, such as satellites, telescopes and particle detectors. The remaining 25% is simulated data based on models of the real world. This quantity of data is roughly the equivalent of 40 million movies. On an annual basis we commit the equivalent of four to six million new movies to long-term storage. This means we process the equivalent of 100,000 movies per day to extract information and create knowledge and value.

AI has the potential to transform the "nature and methods of scientific inquiry"[2], presenting us with the following possibilities to maximise the value of our data:

- AI-enabled data-processing pipelines, to ingest existing and future large experimental and observational datasets to accelerate analysis and prediction, and create deeper and more rapid insight and discoveries
- Digital twinning to enhance the design, manufacture and optimisation of experiments, complex engineering systems and facilities
- Innovative tools for facility control and operation, including instrument automation, anomaly detection, and preventive maintenance scheduling
- Domain-aware AI surrogates and pre-selection models to augment and accelerate high performance computing (HPC) simulation workflows
- Near-real-time coupling of experiment, facility, industrial or biological processes with large-scale simulation and analysis, allowing rapid feedback and optimisation of throughput
- Augmented working via automation of daily tasks (code generation, documentation, knowledge elicitation, explanatory materials)

2. Science in the Age of AI; royalsociety.org/news-resources/projects/science-in-the-age-of-ai

In the fields of particle and nuclear physics, AI already enables a plethora of applications that were inaccessible until very recently. An especially impactful example from particle physics relates to the capability of identifying the origin of sprays of hadrons (jets) in collider experiments. This is referred to as flavour tagging. It is central to a majority of the physics programme at hadron colliders and is critical for some of the highest-profile analyses, such as the search for Higgs boson production. These techniques have been in use for over three decades. Recently paradigm-shifting improvements in performance have been made by exploiting the latest cutting-edge AI techniques such as Graph Neural Networks and transformer-based models at ATLAS. These models have dramatically enhanced the flavour-tagging efficiency in all regions, by factors of two to three. This is equivalent in statistical terms to running the Large Hadron Collider (LHC) for six times longer. As well as advancing performance, the algorithm has come with a host of additional benefits, including enabling greater explainability, the extraction of more precise measurements and a simplified algorithmic approach.

In the field of cosmology, AI will soon take on a foundational role, as it is accepted that the next generation of simulations of the Universe will require AI-enabled acceleration in order to be tractable on even the largest supercomputers.

Going forward, our objective is to use new AI initiatives to advance STFC science and technology, but to ensure that these developments in turn advance AI. Examples of such initiatives are given below.

AI-driven data analysis

AI is essential for analysing extremely large datasets from experimental facilities and observatories. It accelerates data analysis, reducing the time from experiment to results, and produces derived information and measurements with lower statistical and systematic uncertainties.

Edge AI

‘AI at the edge’ technology, where AI is built into experimental and observational facilities and instruments, is crucial for next-generation facilities. This includes the SKA Observatory (SKAO) and the upgraded Diamond Light Source (Diamond) where AI will be a core enabling technology. AI integration is also being considered at the design stage of next-generation infrastructures and instruments across the Central Laser Facility (CLF), ISIS Neutron and Muon Source (ISIS), our Accelerator Science Technology Centre (ASTeC) and the UK Astronomy Technology Centre (UK ATC).

AI’s role extends to real-time control and feedback, including as high-frequency real-time control, which is widespread in STFC applications from space telescopes to accelerators to quantum computers. Model pruning and distillation will be a critical consideration to enable AI on standard computing platforms instead of only HPC, enabling close coupling and integration with our instruments and facilities.

AI for complex workflows and simulations

AI is becoming indispensable in fundamental science. This is particularly the case in disciplines that are underpinned by a combination of large-scale data and simulations, including materials sciences, particle physics, cosmology and life sciences. AI is used to accelerate complex astrophysics simulations, aiding the interpretation of wide-field observational campaigns in terms of our cosmological models. AI is also central to developments in chemistry and life sciences, exemplified by various projects within the UK's exascale software programme ExCALIBUR that STFC is helping to deliver.

Many scientific applications benefit from dedicated, 'smaller' solutions. One example is given by subgrid physics surrogates, in both astrophysics and nuclear physics, to speed up numerical simulations without relying on prohibitive resolution. In weather simulations, intense research is happening on surrogate models of cloud formation that cannot be addressed by brute-force fluid dynamics simulations on even the largest HPC systems.

Foundation models and collective insights for science

Large-language models (LLMs), a sub-set of foundation models, have made rapid inroads in applications for the general public, also supporting scientists working on knowledge distillation or meta-analysis. STFC recognises the broader significance of multi-modal foundation models in scientific and industrial applications. More specifically, STFC remains aligned and committed to using the outputs from across its activities for new insights, the development of new foundation models and use of existing foundation AI models (a superset of LLMs) for advancing science and discoveries. Examples of this include the fine-tuning of foundation models for computational chemistry of wide applicability to the discovery of performant materials for facilities developed and provided by STFC.

To capture the benefits of these initiatives, we will integrate and build on our AI capacity and capabilities. We will create the right conditions, mechanisms and access to compute to deploy AI and ensure it thrives across our core investments. We will sustain a critical mass of AI specialists to co-create, develop and deploy science-driven solutions alongside our AI-skilled domain experts, at scale and pace. This will make STFC's science more productive and open up new lines of enquiry. We will also work to ensure the dynamic exchange of ideas and successes across the organisation to appropriately re-purpose or tune AI solutions.

We also aim to build on our capacity to horizon scan and work with an increasing range of partners in academia, industry and the public sector to create, develop and deploy the AI tools of the future.

Astronomy foundation model at the heart of the ESA Euclid Mission processing pipeline

The ESA Euclid Mission has a UK-developed astronomy foundation model at the heart of its processing pipeline, responsible for providing the first catalogue of 380,000 classified galaxies from what has been described as "the ultimate discovery machine". This foundation model, known as Zoobot, was built on STFC funded PhD work developing deep-learning models for the Galaxy Zoo Project. It was the first foundation model in optical astronomy and has been trained using the largest labelled dataset in astrophysics to date. It was developed alongside the first foundation models for radio astronomy, which were also designed and built by a UK team as part of AI development towards the SKAO project.

Strategic goal 2: Securing the health of the AI ecosystem in collaboration across UKRI and the wider R&I ecosystem

Strategic AI advantage for the UK will be dependent on a healthy AI ecosystem (see Figure 1). Each of the four components of the system – data, compute, people and the development and application of responsible, trustworthy software – need to be optimised and integrated, to advance the application and adoption of AI at scale and pace.

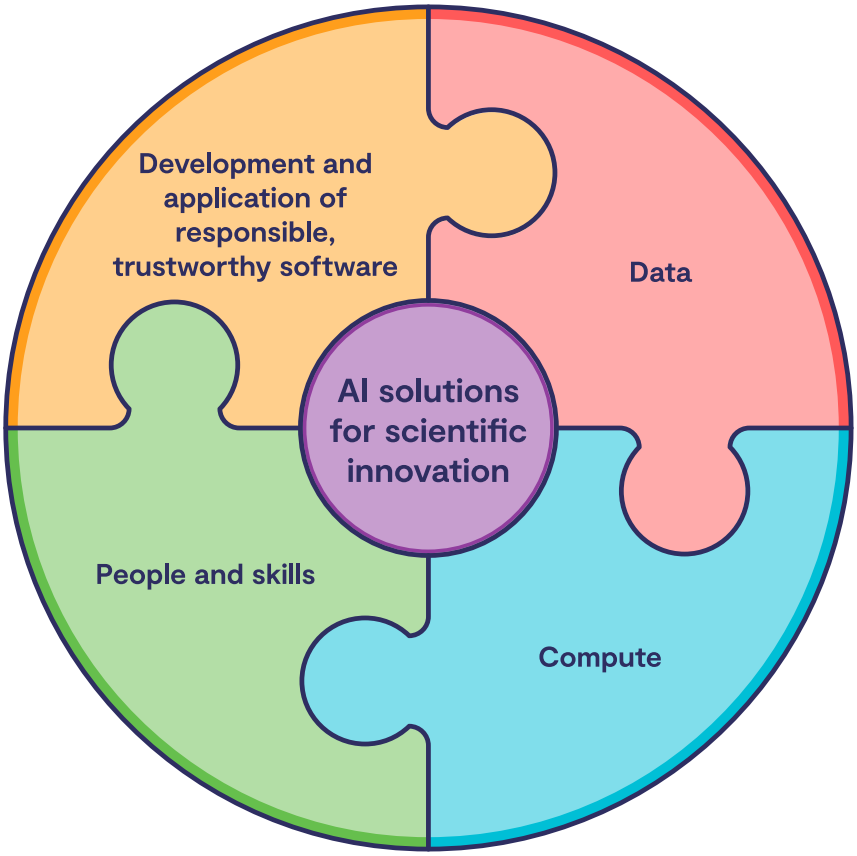


Figure 1: A healthy AI ecosystem

We will therefore ensure that all four components are supported across STFC.

People and skills

Driving the application and adoption of AI across science and the wider economy requires investment to:

- Build and maintain a thriving AI-skilled ecosystem through training and professional development across all career stages and role types, with a focus on equality of opportunity
- Sustain a critical mass of AI specialists and ensure domain experts are appropriately skilled in AI
- Build partnerships to support the development of best practice and enable the flow of expertise across science domains, national boundaries and sectors. This will ensure the dynamic development of new ideas and creative cross-sectorial approaches

Data

Data is critical to AI. The validity and trustworthiness of AI models depend on the quality, quantity and variety of well-curated, meta-data rich data on which they are trained. It is not possible to generate a strategic AI advantage without sophisticated delivery of the whole data-logistics lifecycle, including data infrastructure, data systems, data engineering and AI-ready data curation.

Compute

AI's data-hungry nature is a major driver of ensuring fit-for-purpose compute provision across the R&I ecosystem. The building and training of AI models themselves requires large-scale accelerated compute facilities. Building the potentially autonomous computational workflows requires heterogeneous compute systems on the 'computing continuum' that enable data and compute intensive applications.

As an organisation we must ensure our communities have access to large-scale compute within a national compute ecosystem, tailored to the requirements of the workflows related to AI data processing, model training, inferences and data analysis.

Development and application of responsible, trustworthy software

To remain at the forefront of AI, fundamental AI research must be combined with the development and application of challenge-led tools and technologies, to advance science and drive adoption across missions. A variety of tools, technologies and algorithms are considered AI but they all serve to extract knowledge and understanding from data and drive better, more precise decision-making. The latest technological advances in AI enable the creation of domain-aware, scalable AI, with the capacity to adjust speed, complexity and size depending on the context or challenge.

Whenever we research, develop and serve solutions based on AI and automated decision making, we need to consider issues of safety, explainability, reproducibility, reliability and ethics. The UK AI Safety Resolution and the EU AI Act recommend special scrutiny for narrow or large models that can pose high risk, which directly impacts the development of foundation models for science and AI-aided toolkits with consequences on humans.

Environmental sustainability is also a key consideration when seeking to secure the health of the AI ecosystem.

The AI initiatives discussed in strategic goal 1 have applications across various domains and sectors. They represent only the tips of a vast set of use cases.

Recognising the huge potential impact of AI, work is taking place across the R&I landscape to build strong foundations for an outstanding AI ecosystem. The government's focus to date has been on AI safety, large-scale compute investments and the diffusion of AI across the economy. Key national investments to date include the AI Safety Institute (AISI), Alan Turing Institute, STFC's Hartree National Centre for Digital Innovation (HNCDI), Responsible AI UK and Innovate UK's Bridge AI.

Investment in compute for AI comprises the STFC-mediated AI Research Resource (AIRR) systems at the University of Bristol (Isambard-AI) and the University of Cambridge (Dawn). The UK government announced its intentions to invest up to an extra £1 billion of funding to scale up computing power for AI 20 fold, as well as up to £750 million to build a new national supercomputer for the UK at the University of Edinburgh. This "strengthening Britain's position as an AI-maker and research power" represents a huge step forward in equipping scientists and researchers with the tools to explore and advance AI.

3. www.anl.gov/ai/reference/AI-for-Science-Energy-and-Security-Report-2023

AI is a rapidly moving environment and significant investments are also being made globally.

In the US, the Department of Energy (DoE) National Laboratories, in conjunction with the Office of Science and the National Nuclear Security Administration, are driving the use of AI to "accelerate the progress, and deepen the quality of mission areas spanning science, energy, and security"[4]. Their multi-billion-dollar Frontiers in AI for Science, Security and Technology program has a target to recruit 2,000 people[5].

The EU ETP4HPC Strategic Research Agenda enables the adoption of digital twins and AI-enabled solutions at scale in academia and across the economy using the 'digital transcontinuum' (edge to HPC/Cloud in the data centre, including networks, AI augmentation). Its focus is on:

- **Infrastructure**
- **Software, people and skills with a useful discussion on open source to drive adoption**
- **Quantum for HPC**
- **Environmental sustainability (of infrastructure and optimised software)**
- **EU digital sovereignty**

The AI Opportunities Action Plan will enable the UK government to "capture the opportunities of AI to enhance growth and productivity and create tangible benefits for UK citizens"[6].

As a discipline in its own right, AI can be described as the latest 'big science'. It is computationally intensive, requires interaction with complex systems and a wide variety of data volumes and modalities, and relies upon co-creation between interdisciplinary teams that stretch across international borders.

Because of the data-intensive nature of our science and our role designing, enabling and delivering science at scale, STFC has developed a wealth of experience and capabilities across the AI ecosystem. We work at the intersection of data, compute and software to develop solutions based on the complex, diverse data that we generate and store.

STFC's particle physics, astronomy and nuclear physics (PPAN) science communities have been pioneering the development and deployment of machine learning and AI solutions for decades, to manage and exploit an ever-increasing volume of data and computational power.

4. On the Trail of Exascale and AI conference, 20 – 22 May, 2024
5. www.gov.uk/government/publications/artificial-intelligence-ai-opportunities-action-plan-terms-of-reference/artificial-intelligence-ai-opportunities-action-plan-terms-of-reference
6. www.gov.uk/government/publications/artificial-intelligence-ai-opportunities-action-plan-terms-of-reference/artificial-intelligence-ai-opportunities-action-plan-terms-of-reference

Our National Laboratories play a critical role in driving AI for Scientific Innovation[7]. They sustain a powerful environment comprising over 3,000 in-house theorists, experimentalists, facility designers and operators, biologists, physicists, chemists, mathematicians, technologists, engineers, data scientists and research technical professionals. They work together with academia, industry and the public sector to develop and exploit instruments, facilities, data, compute and software across disciplines to advance discovery and innovation at scale. Use of domain-aware AI solutions is now an expected part of this delivery role.

STFC's Scientific Computing department (SC) supports the computing and AI/ML requirements (including algorithm development and application, data logistics and compute resources) of various STFC facilities and UK scientific communities, including:

- **CLF**
- **ISIS**
- **Diamond**
- **National Quantum Computing Centre**
- **Extreme Photonics Application Centre**
- **SKAO**
- **Fundamental Physics Community**
- **Natural Environment Research Council (NERC) Earth Observation Community**

AI for science is a critical area for SC. The department works with other key enablers and catalysts for the application and adoption of AI funded by or within STFC. These include the Ada Lovelace Centre (ALC), IRIS, digital research infrastructure for STFC science, DiRAC, a HPC facility supporting the STFC Theory Community, GridPP, which manages the UK's involvement in the Worldwide LHC Computing Grid (WLCG), CERN's Large Hadron Collider Computing Grid project, as well as Collaborative Computational Projects (CCPs). Through these collaborations SC supports a broad portfolio of UKRI science communities and STFC facilities to remain at the cutting-edge of AI capabilities.

Across our large-scale facilities, domain experts and AI specialists are developing and deploying AI to make instruments smarter and more reliable for the benefit of our users. The Ada Lovelace Centre, a multidisciplinary scientific computing centre within SC, was established to emphasise the importance of simulations and data analysis across STFC facilities. The centre now coordinates strategically important challenges by drawing on capabilities across SC, to combine domain expertise, data management, mathematics and AI to solve complex problems and deliver solutions and new ways of working.

Image credit: Getty Images

7. Artificial Intelligence in Science, OECD, file:///C:/Users/cdk41992/Downloads/a8d820bd-en.pdf

For example, ALC is combining accelerator physics knowledge with AI-driven surrogate models with mathematical optimisation to tackle challenges in particle beam stability. Through ALC, we are shaping good practice in data stewardship, curation, and logistics for a large number of researchers and innovators using the UK's large experimental facilities. This approach maximises the value of data for the widest set of potential users.

The Hartree Centre is a supercomputing centre focused on exploring the convergence of HPC, AI, data analytics, and quantum computing, supporting their adoption by industry and the public sector. The Hartree Centre is UKRI's largest investment in AI adoption, and is one of the key national delivery organisations[8] driving the digital transformation of the economy. Its flagship Hartree National Centre for Digital Innovation (HNCDI) programme is a public-private partnership that has leveraged £37 million of private investment to date. The core aim of HNCDI is to increase UK productivity and generate economic benefit across UK industry and the public sector through de-risking and increasing the application and adoption of advanced technologies such as HPC, AI, cloud and quantum.

Specialised research and engineering expertise, along with HPC capability, are provided to help companies tackle the complex tasks many face as they strive to gain competitive advantage. At present, there is no other UK-based organisation providing similar AI/HPC support to business. In 2023, the Hartree Centre established three regional SME Engagement Hubs, boosting its ability to support UK SMEs. Through the regional hubs, a focused understanding of local agendas and regional business networks is paired with centralised facilities and highly skilled staff across a wide range of digital technologies. Central to the Hartree Centre's mission is providing solutions to organisations that are medium-to-high level in their technology readiness, though the programme also contains support for businesses at an earlier stage of adoption.

The interdependence between the physical sciences and AI is fundamental to harness the AI potential, due to a variety of factors:

- Our science applications are linked by the scale, complexity, and diversity of data, requiring new and efficient approaches to model generation, training, and validation
- As highlighted throughout this strategy, AI has the potential to be highly impactful in STFC science areas, resulting in a dynamic and highly-motivated international user community willing to invest effort and take controlled risk in return for large potential returns
- The physical sciences provide a 'ground truth anchored' testbed for AI techniques, allowing comparison of outputs with more traditional methods for verification and validation, in a way that is not possible in other more subjective disciplines
- STFC provides researchers with facilitated access to very large-scale compute and data storage, allowing rapid development and testing of new AI techniques
- The continuous flow-through of trained early-career researchers in big-science subjects results in a direct pipeline of new techniques into wider academia and industry

8. UKRI AI Review "Transforming our world with AI. UKRI's role in embracing the opportunity." Jan 2021

STFC science communities, in their wider international context, are able to deliver within their own domains a high pace and scale of innovation combined with a level of scientific rigour and self-oversight that is required to give confidence in large-scale investments. This is the same dynamic that led to the pioneering developments around the world wide web in the early 1990s and in distributed scientific computing in the 2000s.

STFC's initiatives in driving the health of the AI ecosystem, in conjunction with partners across the R&I ecosystem, will ensure we are well-placed to continue to make the changes needed and capture the true potential of AI and create increasing economic impact across the short, medium and longer term.

Strategic goal 3: Enabling organisations across the UK to harness the potential of AI to solve real-world challenges

The health of the R&I system is dependent on academia, National Laboratories, industry, third sector and public sector working together. This partnership model has enabled us to develop a successful approach to oversight, procurement and risk management of the development and delivery of large programmes of science at scale. It is now fundamental to achieving our ambitions around AI advancement.

To realise our vision, we are co-creating and deploying solutions across all components of the AI ecosystem, as shown in Figure 1, in conjunction with our AI specialist and domain expert partners.

Together, we are solving complex challenges and creating far-reaching impact across the economy. We work with research and science organisations, including STFC itself as a consumer of AI, SMEs, public sector and private sector. We also work across domains and sectors which underpin core national missions including science at scale, fusion, healthcare, materials, space, energy, catalysis and business supply chains and logistics.

STFC's deep tech incubation programme supports UK AI startups by providing a structured package of incubation support. The package includes grant funding, R&D vouchers, access to STFC technical facilities, capabilities, and expertise, and expert commercial coaching. The programme helps startups to de-risk their product development, attract private investment and get to market faster. We have a 14-year record of delivering socioeconomic impact across our portfolio, including over £350 million of gross value added to the UK economy and a 25-fold return on investment.

AI for blood cancer diagnostics

Founded in Manchester, Spotlight Pathology integrates AI and image analysis to address the global shortage of trained haematopathologists and improve blood cancer diagnostics.

The diagnosis of lymphomas and other blood cancers is complex and time-consuming, with general pathologists only achieving 40% diagnostic accuracy, leading to delayed and misdirected referrals. Spotlight's AI-powered tool supports pathologists by automating triage, improving accuracy, and accelerating diagnosis times, ultimately benefiting patients through faster and more accurate diagnoses.

Participating in the Digital Business Incubation Centre Programme has given Spotlight Pathology access to STFC Daresbury Laboratory's world-leading expertise. This has enabled them to optimise their computational performance and integrate AI models. It has successfully de-risked product development, secured investment, and strengthened its leadership team, bringing cutting-edge diagnostics to the forefront of cancer care.

We also work with organisations across all four components of the AI ecosystem.

Data

A key highlight includes SC's collaboration with the particle physics community, holding responsibility as a Tier-1 centre for hosting the experimental data from the Large Hadron Collider. They provide round-the-clock support for the computing systems where analyses are performed with the goal of better understanding our Universe.

Provision of Trusted Research Environments in collaboration with other UKRI councils or with industry and public sector is an important element of STFC's role. SC and RAL Space, in partnership with NERC, designed, built and provide ongoing support for an innovative Superdata Cluster for High Performance Data Analytics. This allows researchers, such as those investigating climate change, to utilise and access ever increasing environmental data sets. Known as JASMIN, this was the first system of its type in the world. The Hartree Centre has built large, heterogeneous computing environments on site and in the cloud. Hartree has partnered with the AHRC on their Research Infrastructure for Conservation and Heritage Science project to build a national-scale Virtual Research Environment and secure a petabyte-scale data storage facility for the UK's heritage science communities, hosted at Daresbury Laboratory.

Compute

Through the IRIS Collaboration, the DiRAC HPC Facility, GridPP, SC and the Hartree Centre, we provide large-scale resources to STFC's communities. We are delivering federated resource aggregation to create large-scale digital research infrastructures that specialise in supporting research and innovation at all scales of research data and computing.

STFC has also played a leading role in delivery of AIRR for the UK. AIRR has been designed as a federated but jointly managed national AI-specialised compute facility. It was established to provide world-leading compute capacity to researchers and industry in the UK. It is critical to the UK's international standing. With two AIRR systems now operational, Dawn (at the University of Cambridge) and Isambard-AI (at the University of Bristol), the UK is set to herald a new era of AI discoveries and innovations.

Development and application of responsible, trustworthy software

Across the breadth of STFC's diverse portfolio, AI software solutions are being developed to help solve complex challenges.

Collaborations between the Hartree Centre, IBM and the UK Atomic Energy Authority (UKAEA) are geared towards making scalable fusion energy a reality. AI plays a critical role in the design, simulation and understanding of the fusion process at a scale that has never before been possible, reducing simulation times from days to milliseconds. This bridges STFC's PPAN science programme, the design and operation of complex facilities and knowledge from frontier research into technologies for economic and societal impact.

Another example relates to the complexity of the large datasets collected by experiments such as those at the LHC. This has led to the development and deployment of AI-based approaches for trigger and offline analyses, as outlined in Strategic Goal 1. At the Compact Muon Solenoid (CMS) experiment, one of the two general-purpose detectors at the LHC, UK CMS members have developed an AI algorithm based on convolutional and dense neural networks. This runs in a dedicated hardware device, sifting through the trajectories of the particles produced in proton-proton collisions in just a few hundred billionths of a second.

Determining which of these collisions is interesting enough to keep for later analysis requires state of the art technology. AI accelerated algorithms are much more efficient than existing algorithms for this task and will be a crucial pillar of the entire physics programme for the High-Luminosity LHC phase in the next decade.

Diamond is home to the UK's OpenBind Consortium, bringing together world-leading AI and domain experts. It aims to use automated chemistry and high-throughput X-ray crystallography, to generate the world's largest collection of data, on how drugs interact with proteins, the building blocks of the body. The dataset will potentially be twenty times greater than anything collected over the last fifty years. This will cement the UK's position as a global hub for AI-driven drug discovery.

The consortium, backed with up to £8 million of investment from DSIT's Sovereign AI Unit is a key driver of the AI Opportunities Action Plan. It will close critical data gaps by training new AI models to find potential new drugs and help create better treatments for diseases. It will also help scientists use engineering biology to solve bigger problems, like making enzymes that can break down plastic waste.

OpenBind will offer a core dataset that will drive progress across scientific and technological areas, including predicting molecular structures, designing new molecules and improving research workflows. It will work in tandem with other new methods in order to reduce trial-and-error experimentation, guide better decision-making, and support more efficient exploration of chemical possibilities.

Through the ExCALIBUR-funded BASE-II, STFC is linking AI experts from SC and the University of Leicester, with cosmology domain experts from the DiRAC community at Cambridge and Hertfordshire. Together they are developing advanced AI surrogate models of the evolution of the high-energy outflows from the supermassive black holes at the centres of galaxies. These models have the potential to deliver a step change in the realism of future cosmological simulations. By replacing traditional sub-grid models with surrogates, they can capture all the pertinent physics information.

STFC DiRAC-supported projects have embedded students and Post-Doctoral Research Assistants in the NHS where they have developed AI-based tools, in both administrative and clinical settings, delivering benefits to both patients and efficiency within the NHS.

STFC is pioneering the deployment of automated laboratories with integrated data analysis solutions, to assist with product development. Such a solution has been trialled with simultaneous materials formulation on beamlines at Diamond and ISIS, using AI analysis and interpretation of the multimodal data to determine the next materials composition to be formulated in real time. With forthcoming and proposed instrument and source brightness enhancements of orders of magnitude, such automation schemes promise the most effective exploitation of the investment in large facilities.

High-energy experiments and facilities hosted at STFC and UKAEA, alongside STFC-DiRAC work with the NHS, monitored for its safety consequences in aiding human decision-making, are examples of STFC scientists operating in safety-critical domains. Developing responsible, trustworthy tools is necessary to ensure that standards of academic rigour and ethics can be upheld. The ethical component includes our ambition and commitment, as part of UKRI, to reduce our environmental impact and embed environmental sustainability across all our operations.

People and skills

We are investing to build and maintain a thriving AI-skilled workforce to sustain a critical mass of AI specialists and ensure domain experts are appropriately skilled in AI. We are doing this through training and professional development across all career stages and role types, with a focus on equality of opportunity.

SC provides training through different channels, including ML for science for STFC and STFC-funded facilities and programmes such as the DiRAC training programme. This training extends to the UK academic community via application-based courses offered through CCPs and the Computational Science Centre for Research Communities (CoSeC). It covers AI applications in various branches of science and engineering. SC also runs one of the strategic technical platforms (STPs) called NPRAISE (National Platform for Research Technical Professionals on AI for Science and Engineering). This four-year programme aims to upskill the next generation of research technical professionals (RTPs), including research software engineers, academics, technicians, and engineers. NPRAISE involves full- or part-time employment and secondments at SC for up to two years, allowing RTPs in the UK community to gain practical skills alongside experts at the intersection of AI, science, and engineering. The programme also provides unprecedented levels of exposure to the research and development ecosystem within STFC.

The Hartree Centre is recognised as a trusted partner to provide a diversity of training and workforce upskilling throughout sectors. They are delivering BridgeAI together with Innovate UK and other agencies, participating in centres for doctoral training and student placements and offering general and bespoke training in AI, HPC, and quantum computing through the HNCDI Explain strand. A noteworthy example is NSG Pilkington, whose employees (ranging from materials scientists to legal) have enrolled in 61 Explain courses, helping to shape their internal approach to the development of new manufacturing projects.

Through the STFC Skills Centre, we provide a range of early careers computer science opportunities from level three apprenticeships to graduate placements, which, alongside our Centres for Doctoral Training (CDTs) in Data Intensive Science will be critical to drive our pipeline of AI specialists.

Nurturing skills for future data scientists

Through our investments in Centres for Doctoral Training (CDTs) in Data Intensive Science, we are nurturing the capabilities, talents and skills needed to address the data challenges presented by our science and applying them for broader economic and societal benefit. Since the establishment of the first CDTs in 2017, we have funded over 200 doctoral students to carry out original research projects that bring together big data skills with expertise in STFC's core science areas and apply their learning to real-world problems through internships outside of academia. The achievements of our CDT interns range from boosting the competitive advantage of a UK online fashion retailer by applying recent advances in AI to enhance customer care, to providing insights on the spread of COVID-19 to national decision makers during the pandemic through the application of machine learning based prediction models to a national sewage testing programme.

Collaboration between people across domain and international borders is fundamental to successfully advancing AI. Our expertise and track record in AI is driving international collaboration in fundamental, applied and scalable AI research, strengthening the UK's position. There are numerous international collaborations that underpin STFC's AI agenda, including AI for Science and AI for Nuclear Fusion programmes with the DoE laboratories in the US, and AI for Science programmes with research organisations in India. STFC is also a founding member of the Trillion Parameter Consortium, an international consortium established by the Argonne National Laboratory to advance the application of large-scale AI models in science. European collaborations cover both AI and HPC (e.g. ETP4HPC). These collaborations remain central to our innovation and ensure that we stay at the cutting-edge of AI.

For STFC to continue to be valued as an equal or competitive partner in these collaborations, across the four components of the ecosystem, it is essential we are skilful in applying AI to science. We must also be innovative in developing and applying next-generation AI technologies for science and sustaining an AI-skilled R&I system. We need to develop appropriate mechanisms to nurture the interdependence of our skills, expertise and capabilities and ensure they can be seamlessly translated and exchanged across and between STFC and beyond. We must also lead and support national advances across the AI ecosystem.

Image credit: Getty Images

Our strategic objectives: driving AI for Scientific Innovation

To realise our vision, we will drive the application, adoption and integration of AI as a core enabler and driver of STFC's science and innovation endeavours. We will promote collaboration and partnership, increase productivity, and support new breakthroughs in our core disciplines. We will do this whilst maintaining and expanding our role as a provider of solutions and skills to address complex problems across sectors.

We have identified five core strategic objectives, supported by an additional cross-cutting objective of partnerships. These priorities will guide STFC's decision-making over the next five years.

Strategic objective 1: AI skills

Our ambition is to consolidate and enhance our activity across the skills and training spectrum and sustain a thriving, equal and inclusive pipeline of AI-literate domain experts and AI/data specialists. This will apply across career stages, in STFC and beyond, generating a deep organisational understanding of the potential of AI, to be able to make informed choices and benefit from it.

Over the next five years we will achieve success through:

- Identifying and delivering the most appropriate types of skills, training and outreach programmes and opportunities to nurture and develop AI and data expertise
- Nurturing the next generation of AI specialists by providing opportunities to collaborate across STFC and gain exposure to the development and deployment of practical, real-world AI solutions
- Expanding the goals and reach of our AI training programmes for domain experts to enable communities to make informed decisions about adopting AI across their disciplines
- Working with partners to identify mechanisms to exchange knowledge and skills across the economy
- Integrating, consolidating and enhancing STFC's AI expertise and cohering a critical mass of AI specialists. We aim to ensure all minds are focused on the shared goals of AI for Scientific Innovation to enable the necessary step change in the pace of delivery

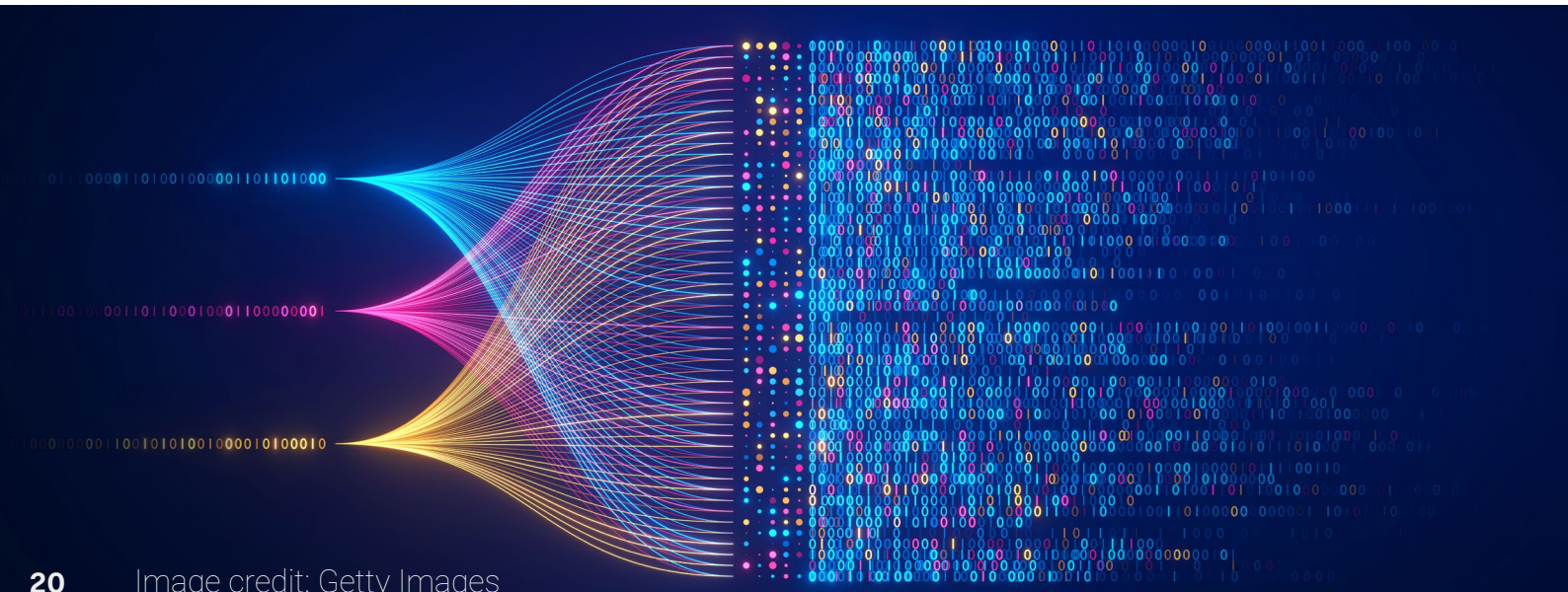


Image credit: Getty Images

Strategic objective 2: AI-ready data

Our ambition is to ensure we can derive value from and develop solutions based on a variety of modalities, volumes and complexities of data.

Over the next five years we will achieve success through:

- Identifying and closing any gaps in our data standards policy, adhering to FAIR principles
- Ensuring that management and oversight of our own and others' data assets on a day-to-day basis – and the data stewardship policies that enable this – are fit for purpose to deliver meta-data rich, AI-ready data
- Ensuring the safe, secure flow of data within the R&I ecosystem to enable the development of AI tools and techniques
- Working with data owners and providers to capture opportunities to maximise the value of data
- Creating a data logistics centre of excellence within STFC to work across our core areas
- Working across UKRI and government and applying our data logistics expertise to the design and delivery of national data capabilities to support new, more productive science

Strategic objective 3: AI challenge-led technical solutions

Our ambition is to maximise the benefits of our investments as well as leverage new resources by investing in and directing our AI R&D activities for research acceleration, insight and facility optimisation and address the most complex science, technology, and engineering challenges.

Over the next five years, we will achieve success through:

- Developing challenge-led AI solutions based on a cycle of 'research, develop, deploy'
- Creating and deploying AI solutions tailored to the digital maturity of the domain context. The solutions will be based on a range of opportunities. These include algorithmic research, edge-AI, and services to support high-bandwidth, low latency data processing; novel model architecture development; adapting off the shelf APIs; and orchestrating workflows with novel or off-the-shelf components
- Identifying cases across our core investments where the application of AI is likely to be most impactful
- Driving innovation in AI by leveraging the depth of expertise across the communities supported by STFC
- Creating ways of working that promote horizon scanning, thought leadership, agility and internal integration and collaboration
- Leading the exploration of the convergence of technologies, for example between quantum technologies and AI, future-proofing our capacity to solve our most complex science challenges

Strategic objective 4: Mission-led AI solutions

Our ambition is to reflect our key enabling role within UKRI and the wider UK R&I ecosystem, leveraging our experience and skills. We will translate our tools, approaches, and solutions across the broader R&I landscape, to solve complex challenges that underpin core national missions and to provide bespoke facilities and resource to support UK organisations to adopt AI and grow 'UK PLC'. We will work with our network of AI partners on government priorities where we believe we can have maximum impact, namely driving economic growth, advanced manufacturing, clean energy, life sciences and future-proofing our healthcare system.

Over the next five years we will achieve success through:

- Working with our partners to co-create and drive delivery of AI solutions for complex challenges in materials, fusion, space, catalysis, climate and environmental science, healthcare diagnostics and therapeutics, public planning, policy and administration
- Driving AI projects which combine and exploit the rich experimental capabilities and data from across STFC for new knowledge
- Defining the boundaries of AI research leadership with our partners
- Working with regional government, mayoral authorities and public services, as well as academic and industry leaders to create thriving AI ecosystems on and beyond our campuses

Image credit: Getty Images

Strategic objective 5: Responsible AI innovation

Our ambition is to develop excellence in responsible AI innovation. We will work in collaboration across UKRI and with partners across the R&I ecosystem to maximise the benefits of AI for Scientific Innovation. We will seek to minimise unintended consequences, particularly with respect to academic rigour and ethics, explainability of AI and environmental sustainability.

Over the next five years, we will achieve success through:

- Embedding responsibility and trustworthiness alongside technology development and innovation. This will serve to make models safer, explainable and at the same time enhance their performance
- Embedding explainability, reproducibility, safety and sustainability in our skills programmes and decision-making
- Establishing and adopting a clear set of responsible AI innovation principles
- Identifying the most appropriate governance and oversight channels for our AI innovation
- Understanding the cost/benefit of applying AI to maximise productivity taking account of both monetary and environmental costs
- Working across the R&I system to engage in the development of responsible, high-efficiency / low-energy AI tools
- Working with industry partners to deploy testbeds for the use of next generation devices and platforms and identifying those that can deliver more energy-efficient AI solutions. We will use our existing expertise to ensure appropriate convergence of these technologies with existing solutions and promote the use of novel technologies for AI

Cross-cutting objective: Partnerships

The principle of working across the system in partnership crosscuts each of our core strategic themes. Recognising that we cannot work in isolation to achieve our ambitions, partnerships are integral to the success of this strategy. As per the biggest science and innovation successes of the last fifty years, leadership in AI will necessarily be a shared endeavour across national and international borders, as well as across domains and disciplines.

We will collaborate with partners globally to shape the future of AI leadership and ensure the seamless exchange of knowledge, skills, capacity and expertise. We anticipate strengthening partnerships and collaborations across UKRI and within multiple sectors including publicly funded R&I organisations, academia, public sector research establishments, third sector, public sector and industry.

Strategic enablers

The success of this strategy is dependent on four key enablers.

Enabler 1: People and governance

To enable delivery of this strategy, we must ensure we have the appropriate leadership, governance and project management structures and processes in place, providing coherence and integration to our AI activity. We must embed AI skills and capability into our strategic workforce planning and ensure our recruitment and retention policies reflect the needs of our AI ambitions.

Enabler 2: Delivery models and mechanisms

To create transformation through the application of AI we will identify the most appropriate strategic funding models across our core programme and national labs and facilities. We must also identify and create a mechanism to act as the intersection between people (AI, data and domain experts), data and compute to catalyse the co-creation of AI tools and solutions for expert and novice users as appropriate and drive the advancement of AI across our science and technology.

Enabler 3: Access to compute

Recent investment in compute is a huge step forward, and STFC aims to play a central role in the design and delivery of a UK-wide federated compute system supporting AI applications and data access. A variety of platforms will be needed across a range of requirements, from fundamental physicists through to industry, and we will continue to work across UKRI to take a leading role in the design and delivery of a world-class federated compute infrastructure in the UK.

Enabler 4: Funding

Whilst lowering barriers to advancement is partly cultural and organisational, funding for data logistics, people and software development will drive truly transformative impact. We will work collaboratively across UKRI and DSIT to identify, leverage and maximise strategic funding opportunities and exploit STFC's unique position in the publicly funded R&I ecosystem.

Image credit: Getty Images

Delivery

We recognise the ambition of our strategy. Successful implementation will require a detailed delivery plan, including strategic prioritisation of objectives against available funding. This will be produced within six months of the adoption of this strategy.

Conclusion

AI is a transformational technology, and the rapid progress in innovation, adoption, and public awareness means that we are at a critical juncture. Deeper and more rapid adoption of AI by STFC has the potential to

- Transform and accelerate our science
- Ensure STFC and the UK maintain their position of global scientific leadership
- Create lasting economic and societal impact

We look forward to working with partners across the R&I system to drive AI for Scientific Innovation at scale and pace.

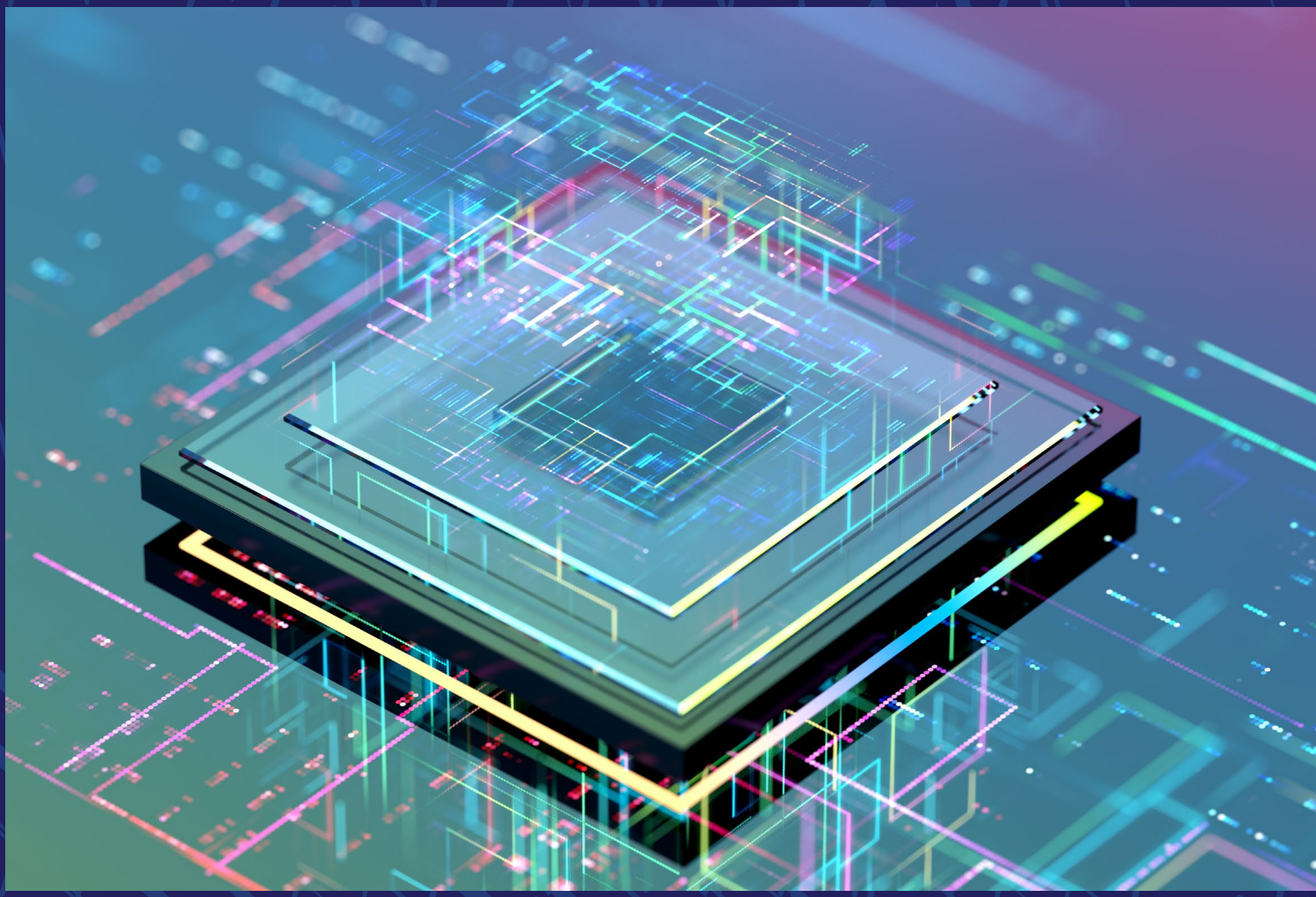


Image credit: Getty Images

Produced by Valerie Farr, STFC.

In conjunction with Mark Wilkinson, Director, DiRAC and Adriano Agnello, Principal AI Researcher, Hartree Centre, STFC.

We'd like to thank all our colleagues and contributors across STFC, UKRI and beyond whose insight and expertise have been instrumental in shaping the direction and content of this strategy.

Special thanks must go to:

- Rob Akers, Director of Computing Programmes, UKAEA
- Vassil Alexandrov, Chief Science Officer, Hartree Centre, STFC
- Carla Andreani, Professor of Applied Physics, University of Rome Tor Vergata and member of STFC Science Board Facilities and Labs
- Grahame Blair, Executive Director, Programmes Directorate, STFC
- Monica D'Onofrio, Professor of Physics, University of Liverpool and member of STFC Science Board PPAN
- Dave Dunning, Senior Scientist, ASTeC, STFC
- Tom Griffin, Director, SC, STFC
- Jon Hays, Head of Particle Physics Research Centre, Professor of Physics, Queen Mary, UoL
- Barbara Montanari, Head of Computer Science and Engineering, SC, STFC
- Massimo Noro, Business Development Director, STFC
- Eleanor Okamoto, Stakeholder Engagement Coordinator, STFC
- Toby Perring, Senior Scientist, ISIS, STFC
- Robin Pinning, Chief Technical Officer, Hartree Centre, STFC
- Cinzia Porcedda, Head, DRI and Astronomy Programmes, Programmes Directorate, STFC
- Paul Quinn, Director, Ada Lovelace Centre, STFC
- Dan Rolfe, Senior Data Scientist, CLF, STFC
- Kate Royse, Director, Hartree Centre, STFC
- Tim Snow, Principal Software Scientist, Diamond Light Source
- Jeyan Thiyagalingam, Head of SciML, SC, STFC

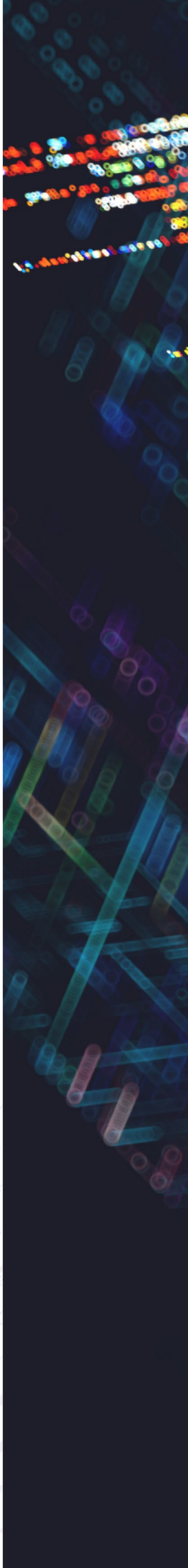


Image credit: Getty Images

Science and
Technology
Facilities Council

ukri.org/councils/stfc

linkedin.com/company/stfc