April 2026

UKRI Digital Research Infrastructure Programme

Pilot project Completion Report

1. Project Details

Project Name:	Enabling AI in High Energy Physics
DRI Project no.:	51
Lead organisation:	University College London
Date of report completed:	30/06/2026
Contact Details: *[In case any clarifications / follow up are needed]*	Tim Scanlon timothy.scanlon@ucl.ac.uk, Monica D'Onofrio <onofrio@liverpool.ac.uk> Davide Costanzo <d.costanzo@sheffield.ac.uk> Alex Tapper <a.tapper@imperial.ac.uk>

2. Purpose and objectives

Provide a brief overview of the project, outlining the context and rationale for the pilot and confirming the original aims and objectives.

High Energy Physics (HEP) has been developing and applying AI and machine learning methods for around four decades. However, it is now entering a new phase in which modern AI can transform the design, operation and analysis of major experiments, and the expertise developed in HEP can be applied to a wide array of applications beyond. HEP is also one of the few research domains with the data scale, increasing numbers of open datasets, software maturity, statistical discipline and computing infrastructure needed to test AI methods under genuinely demanding scientific conditions. This pilot therefore aimed to support both AI for HEP (AI4HEP) and HEP for AI (HEP4AI): using AI to accelerate discovery in particle physics, while using the unique strengths of HEP to advance trustworthy, efficient and scientifically grounded AI for the wider UK research ecosystem.

The central purpose of the pilot was to create the foundations of a coordinated UK expert community for AI in HEP. The project brought together experimentalists, theorists, software specialists, research software engineers, computing infrastructure experts and applied AI researchers. The aim was to move beyond isolated activity within individual experiments and establish a route for shared tools (with flagship demonstrators), reusable datasets, common training material and stronger routes for knowledge exchange. This is essential because many of the AI challenges faced by different HEP experiments are technically similar, even when the scientific questions differ. Shared approaches can reduce duplicated effort, accelerate adoption, and allow researchers at smaller institutions or working on smaller experiments to benefit from the most advanced AI developments.

The project also aimed to lay the groundwork for HEP4AI as a strategic UK capability. HEP experiments produce exceptionally large (LHC detectors generate about 1 PB/s of collision data) and complex (the next generation of LHC detectors will capture information from billions of sensors every collision) datasets, require robust real time decision making, and demand high levels of reliability, validation and uncertainty quantification. These features make HEP an ideal environment for testing foundation model ideas, scaling laws, efficient AI, responsible AI, scientific reasoning systems and AI methods for distributed research infrastructure. The pilot therefore sought to establish the first practical building blocks for this wider ambition: open source cross experimental tools, high quality training datasets, shared software frameworks, fast AI for real time systems, AI assisted operations monitoring, demonstration of testing fundamental AI scaling laws, and clearer routes to common computing and software resources. Together, these activities were intended to show that HEP can both benefit from AI and provide the UK with a distinctive, scientifically rigorous environment for developing and validating the next generation of AI methods.

Finally, the pilot aimed to test whether these emerging activities could form the basis of a more sustained community structure, such as a Centre of Community Excellence for AI in HEP. Over a short period, the project set out to identify the right people, demonstrate the scale of demand, create shared routes for communication, support rapid cross experimental projects, and gather evidence for a longer-term programme. **The pilot was designed as a seed and coordination investment, not as a complete delivery programme for the full AI4HEP agenda.** Its purpose was to establish the groundwork: showing that the UK HEP community can organise around AI, deliver useful shared outputs quickly, train highly skilled applied AI researchers, and provide STFC and UKRI with a credible platform for future investment in both AI for HEP and HEP for AI.

3. Key findings and next steps

Summarise what the pilot delivered and its key outcomes or findings. Briefly reflect on how effective the approaches trialled were in practice, including how they addressed the original challenge the pilot was designed to tackle. Where appropriate, outline any next steps, either as recommendations for the DRI programme or as plans the project team or council intends to take forward independently.

Overall findings

The pilot demonstrated that a small and agile DRI investment can unlock a large amount of high-quality activity when it is targeted at an organised and technically mature community. Several of the activities built on goals and technical directions that the UK HEP AI community had already identified, via self-organised bottom-up activity, before the pilot began. The value of the DRI investment was to bring these strands together, provide targeted support where it could make a rapid difference, and turn existing momentum into coordinated community outputs, demonstrators and evidence for future investment.

The response from the UK HEP community was immediate and strong. A rapid one week call for short-time individual projects (referred to as rapid response projects, RRPs) produced 17 submissions across AI infrastructure, real time AI, reconstruction, monitoring, datasets and analysis, submitted for 4-6 weeks worth of funding. This showed that there is substantial latent demand, a strong pool of ideas, both deep and broad expertise, and a community ready to deliver when modest seed funding is available. Three RRPs were selected to go ahead in addition to the core projects (CPs) defined at proposal stage by co-lead institutes (Liverpool, Manchester, Sheffield and Imperial College London).

The project also confirmed that HEP is a strategic asset for UKRI AI ambitions. It is not only a user of AI. It is a demanding scientific environment that is highly relevant to UKRI priorities in AI for science, responsible AI, efficient AI, data infrastructure, skills and research software.

The outputs described below include a mixture of activities directly funded by the pilot, work delivered by researchers and teams supported through the pilot, and closely related developments that the pilot helped to coordinate, accelerate, connect or expose through the AI4HEP community. This reflects the purpose of the investment as seed funding: to mobilise an existing technically mature community, strengthen coordination, and turn ongoing expertise into visible, shared progress.

Community building and coordination

The pilot helped establish a more coherent UK AI in HEP community, bringing together researchers from multiple experiments and institutions, including collider physics, neutrino physics, theory, computing, software and infrastructure. These meetings shifted the discussion from identifying opportunities to organising delivery.

Key community outputs included:

1. Two national workshops organised focused on AI in HEP [Imperial ~50 attendees, Online Progress Update ~20 attendees].
2. A rapid response project call that generated 17 proposals on a very short timescale, with 3 awarded in addition to the ones initially included by the proponents.
3. Built stronger connections between HEP researchers from different experiments, RAL Scientific Computing Department, STFC infrastructure teams and UKRI AI strategy discussions.
4. Development of community communication routes, including a website (webpage), mailing list (e-group), Mattermost channel (mattermost), shared resources/github actively being populated (repository), more regular coordination and a more formal structure.
5. A clearer shared vision for moving from AI for HEP towards HEP4AI, where HEP contributes datasets, expertise and validation environments to national AI development.

Open datasets and AI training resources

The pilot supported the creation and wider use of high-quality datasets that can be used to develop and benchmark AI methods. This is a major contribution because many AI researchers outside a specific experiment cannot easily access realistic scientific training data. Public or reusable datasets lower barriers to entry, make results easier to compare, and allow researchers from AI, computing and other fields to work on HEP relevant challenges.

Deliverables included:

1. Contribution to the release of one of the largest official HEP machine learning datasets produced to date: a flavour tagging dataset containing several hundred million jets, together with the associated open-source training software needed for AI method

development, benchmarking and reproducibility studies [*in final ATLAS approval*]. This expands and improves upon the ~50M jets already released [dataset]

2. Development of plans for a future jet dataset at a much larger scale, using between roughly 500 billion and 1 trillion jets already available within ATLAS. This would provide a scientific AI training dataset at a scale comparable to the largest AI datasets used internationally, and would enable HEP to test AI scaling behaviour in a rigorous scientific setting [dataset moving towards internal production].
3. Contributions to realistic collider datasets designed to provide low level detector information for AI development, including ColliderML style datasets for high occupancy LHC and future collider conditions paper [dataset, enhanced versions in preparation].
4. Development of the Collide2V dataset for event level representation learning. This includes more than 50 high pileup physics processes and close to one billion events, with both full granularity and trigger granularity views of the same events. A 1 percent prerelease is already available, with the full dataset expected to be of order a few hundred terabytes [dataset].
5. Studies using the above datasets as the basis for future work on representation learning, trigger emulation, scaling law studies and foundation model style approaches for HEP, with the large jet-tagger models already highlighted as such models.
6. Early work and discussions on making other experimental datasets available for AI development beyond collider physics, including opportunities in intensity frontier and neutrino experiments.

Cutting edge reconstruction and particle identification

The pilot supported AI methods that directly improve the scientific capability of experiments. These methods help reconstruct particles more accurately, identify rare signatures more effectively and extract more information from existing detector data. The results are significant because they are not only incremental software improvements. In several cases, they show step changes in performance, speed or reconstruction quality that could translate into improved physics reach, better use of existing detector data and more efficient future experiments.

Deliverables included:

1. Supporting further development and use of advanced flavour tagging algorithms, including GN2, GN3 and GNX style taggers [paper]. These algorithms identify the type of heavy flavour particle or quark that produced a jet, which is essential for key measurements and searches at the LHC. Newer methods have delivered around a further factor of three improvement in performance compared with older taggers and their deployment in triggering systems on the experiments.
2. Contribution towards scaling studies for next generation flavour tagging [public note], showing that much larger training datasets and models could deliver up to a further factor of around 50 in performance. This is a major strategic result because it suggests that HEP can test AI scaling behaviour in a rigorous scientific environment, using well defined tasks, controlled datasets and physics meaningful performance metrics.
3. Extension of advanced graph and transformer-based tagging methods towards tau reconstruction [public note]. Tau leptons are central to Higgs, electroweak and beyond Standard Model physics, but are difficult to identify cleanly in busy collision environments. Applying these methods to tau reconstruction offers a route to better signal efficiency, stronger background rejection and improved sensitivity in important physics channels.

4. Further development of MaskFormer style tracking algorithms for reconstructing particle trajectories from detector hits [paper]. These approaches have demonstrated very large speed improvements, with inference times of around 100 milliseconds and approximately 100 times faster execution than standard CPU based tracking in relevant tests, promising significant reduction to one of the most compute heavy reconstruction algorithms. Further improvements recently demonstrated [paper].
5. AI enhanced vertex reconstruction for liquid argon neutrino detectors, improving the identification of the point where a neutrino interaction begins. In the studies discussed, the fraction of vertices reconstructed within 1 centimetre improved from approximately 60 percent to approximately 80 percent, which is a substantial gain for event reconstruction in experiments such as DUNE and SBND.
6. AI enhanced 2D to 3D pattern recognition for liquid argon time projection chambers, improving the reconstruction of electrons, photons and low energy particles in complex neutrino interactions. This is particularly important for high multiplicity events and for physics signatures where low energy activity is scientifically important.

Cross experimental AI tools and software

A major success of the pilot was the development and wider testing of AI tools that can work across experiments. Cross experimental development is especially valuable in HEP because many experiments face similar technical challenges: reconstructing tracks and vertices, identifying particle interactions, separating signal from background, combining detector information and validating models under demanding conditions. Developing these tools collaboratively reduces duplicated effort, improves software quality, allows methods to be benchmarked in different detector environments, and helps smaller experiments benefit from methods originally developed for larger collaborations. It also makes the UK community more coherent, since researchers can build common software, datasets and expertise rather than working in isolated experiment specific silos.

Deliverables included:

1. Tests of the MaskFormer tracking algorithm (developed on ATLAS Opendata) in the Mu3e experiment, where the detector layout and event environment differ significantly from the LHC, and also on the LHCb experiment, showing promising out of the box performance in a challenging flavour physics environment. This is an important validation of whether modern AI tracking methods can generalise beyond their original experimental context and a demonstrator of cross-experimental AI models.
2. Support for HEP specific model training frameworks such as SALT [repository] and MaskFormers, enabling wider use beyond the original development teams and helping new users train, compare and deploy modern architectures more easily.
3. Targeted integration of AI into Pandora reconstruction workflows, supporting future cross experimental use in DUNE, SBND and other liquid argon detector programmes.
4. Development of LLM based documentation and workflow tools, including ChAtlas and related retrieval augmented generation approaches that have been forked or adapted by other experiments. These tools help researchers find technical information, navigate large documentation systems, support operations knowledge management and reduce barriers for new users [repository].
5. Supported the development of AgentRivet, an automated system for producing Rivet analysis routines from journal publications, demonstrating how commercial large language models can generate syntactically competent routines with reasonable

physics fidelity while also exposing the validation and oversight needed before such tools can be trusted in production HEP workflows [paper].

Fast AI and efficient real time decision making

The pilot supported FastML, which develops AI methods that can run in real time on specialised hardware [roadmap]. This is one of the strongest examples of HEP4AI within the pilot. Future experiments must make intelligent decisions within microseconds, using incomplete detector information, under strict constraints on latency, power, memory and reliability. That makes HEP a uniquely demanding testbed for cutting edge AI: models must be accurate, compressed, robust, verifiable and deployable on real hardware, not only performant in offline benchmarks. Progress in this area has benefits well beyond HEP, including autonomous instruments, accelerator control, fusion, robotics, edge AI, medical imaging, safety-critical systems and other environments where AI must operate safely and efficiently at very high speed.

Deliverables included:

1. Advanced the UK FastML ecosystem through work on FPGA-targeted AI tools, including improved support for Altera devices, developed in collaboration with Altera UK, and contributions to hls4ml [paper], HGQ and related optimisation workflows.
2. Further development and testing of compact transformer and graph neural network models for deployment on FPGA hardware, in collaboration with Altera UK, including transformer models designed for sub-microsecond inference in trigger and real-time decision-making environments.
3. Development of graph neural network optimisation methods for FPGA deployment.
4. The JediLinear approach for efficient graph neural networks on hardware, including award winning work in field programmable technologies.
5. Development of training material and tutorials for students and early career researchers. Ran a tutorial for people in the UK, the day after the Imperial workshop [agenda ~30 attendees].
6. Exploration of foundation model ideas for future trigger systems, including approaches where large offline models are distilled into smaller real time components.
7. Explored PHAZE-1, an early-stage approach to verifiable low-latency AI inference for future trigger systems, using early-exit ideas and proof-based methods, including zero-knowledge proofs, to investigate how fast AI decisions could be checked for computational integrity and reliable deployment.

AI for computing infrastructure and operations

The pilot also showed that AI can help operate the digital research infrastructure itself. HEP relies on large, distributed computing systems, which generate many alerts and require expert human intervention. AI can help identify problems earlier, triage issues, reduce operational load and improve reliability. The pilot also highlighted Inference as a Service as an important route for AI-enabled HEP computing, allowing grid workflows to use centrally hosted GPU inference services and thereby supporting larger models without placing the full memory and hardware burden on individual jobs.

Deliverables included:

1. Developed an initial demonstrator direction for using AI to support distributed computing operations, including connections to local monitoring systems such as Zabbix and exploration of Model Context Protocol tools to let AI systems query operational data.
2. Investigated AI methods for alert filtering, issue triage and solution identification, with early work on links to ticketing and experiment monitoring systems, alongside consideration of safe human oversight for future semi-automated responses.
3. Studied GPU-based Inference as a Service at a WLCG Tier 2 site, identifying current bottlenecks for smaller models while establishing a clear route towards larger future models that cannot be run efficiently within standard grid jobs.

Skills, training and wider UK benefits

The pilot strengthened the UK skills pipeline in applied AI, one of the most important wider benefits of such investment. HEP researchers are trained to work with complex data, advanced statistics, large software systems, distributed computing and international collaboration. Adding modern AI training to this skill base produces people who are exceptionally well placed to contribute to UK science, industry, public sector AI deployment and AI sovereign capability.

Deliverables and benefits included:

1. Training opportunities through workshops, tutorials and funded rapid response projects.
2. Hands on experience for early career researchers in applying AI to real scientific and infrastructure problems.
3. Development of researchers who understand both AI methods and the reliability requirements of high consequence scientific systems.
4. Stronger links between physicists, research software engineers, computing experts and AI specialists.
5. Reusable training material in areas such as FastML, AI reconstruction, dataset use and model deployment [tutorial].
6. A stronger base for future collaboration with national AI centres, RAL infrastructure, data intensive science CDTs, the Alan Turing Institute, Hartree, SciML and industry activities.
7. A contribution to the national need for applied AI experts who can build, validate and operate AI systems responsibly.

Next steps

The AI4HEP pilot has demonstrated the value of a coordinated national community bringing together AI researchers, High Energy Physics scientists, software engineers and research computing experts. The next step is to build on this momentum by establishing a sustainable community structure that can play a leading role in the emerging UKRI DRI ecosystem.

In particular, following the announced plans, the community could work towards establishing a Community Centre of Excellence (CCE) for AI4HEP/HEP4AI, providing a national focal point for AI-enabled scientific computing. Such a centre would coordinate expertise across experiments and institutions, support access to national AI and compute infrastructure, develop shared software and datasets, and provide a natural interface to UKRI, STFC and the wider research community. It would also position HEP to contribute to future UKRI Grand Challenge programmes and other strategic national initiatives.

Recommended next steps include:

1. Establish a formal long-term UK AI in HEP community structure with long-term STFC support.
2. Develop a proposal for a Centre of Community Excellence that connects HEP experiments, theory, computing, software, RSEs and national AI infrastructure, aligned with the UKRI DRI strategy.
3. Continue to develop, curate and test cutting edge scientific AI datasets for HEP, including datasets that are open, well documented, reusable across experiments and large enough to support modern AI development.
4. Use HEP data and workflows as a testbed for studying AI behaviour at scale, including scaling laws, foundation model approaches, uncertainty quantification, robustness, validation and efficient deployment in demanding scientific environments.
5. Develop HEP4AI as a strategic UKRI theme, positioning HEP as a testbed for trustworthy scientific AI.
6. Strengthen partnerships with industry through joint research activities, an Industry Advisory Board, showcase events and collaborative technology demonstrators, ensuring that developments are informed by industrial best practice while maximising opportunities for knowledge exchange and impact.
7. Continue to build stronger links with the wider UK AI initiatives such as data intensive science CDTs, FastML Foundation, RAL Scientific Computing Department, national AI research infrastructure and UKRI AI for science programmes.
8. Support long term career paths for applied AI experts, research software engineers and technical specialists, recognising their critical role in translating AI research into production scientific capability.