Building an AI-native Research Ecosystem for Experimental Particle Physics: A Community Vision

February 10, 2026 - Version 0.80 (draft)

Abstract: Experimental particle physics seeks to understand the universe by probing its fundamental particles and forces and exploring how they govern the large-scale processes that shape cosmic evolution. This whitepaper presents a vision for how Artificial Intelligence (AI) can accelerate discovery in this field. We outline grand challenges that must be addressed to enable transformative breakthroughs and describe how current and planned experimental facilities can implement this vision to advance our understanding of the vast and complex physical world from the smallest to the largest scales. We show how facilities currently under construction, such as the HL-LHC and DUNE, can both benefit from and serve as proving grounds for this vision, while also enabling a longer-term goal for how future experiments— like FCC-ee at CERN, IceCube-Gen2, a Muon Collider in the U.S., and smaller to mid-scale projects—can be fully AI-native. We describe how a truly national-scale collaboration, jointly managed across large funding sources, and involving both DOE laboratories and universities, can make this happen.

Audience: The goal of this whitepaper is to highlight the emerging opportunities and existing gaps for the broader particle physics community, to inform funding agencies in the event new resources become available, and to highlight to policymakers the innovative contributions that experimental particle physics can bring to the field. This document is not intended as a comprehensive proposal for all necessary activities, nor as an exhaustive review of ongoing R&D.

This report has been produced in collaboration with the Coordinating Panel for Software and Computing (CPSC) which is hosted by the Division of Particles and Fields (DPF) of the American Physical Society (APS).

Contents

Executive Summary

Experimental particle physics addresses some of the most fundamental questions about the universe through facilities that are among the largest, most complex, and ambitious scientific endeavors ever constructed. Across collider, neutrino, cosmic, and rare-event experiments, these facilities function as *massive and continuous data generators*, producing petabytes of rich, structured, curated data annually, while discarding a majority of the raw information due to bandwidth, storage, or latency constraints. The scale, complexity, and structure of these datasets align with the strengths of modern Artificial Intelligence (AI): high-dimensional pattern recognition, rare-signal inference, low-latency decision making, and the orchestration of complex systems spanning hardware, software, and human expertise. AI can play a transformative role by enabling experiments to extract and retain more information from data, extending the discovery potential, and reducing the time from data-taking to discovery. It can also improve the efficiency and sustainability of long-running facilities and increase sensitivity to subtle or unexpected phenomena. Now is a pivotal moment: experiments currently in operation or under construction will define the scientific output of particle physics for the next several decades, while unprecedented national investments in AI, advanced computing, and workforce development create a rare opportunity to couple our scientific challenges with foundational AI research. This whitepaper presents a community vision and an actionable plan to seize this moment.

Our vision is to embed AI end-to-end across the experimental lifecycle, from the co-design of accelerators and detectors to intelligent sensing, data acquisition, autonomous operations and calibration, and accelerated analysis for discovery. In this "AI-Native" paradigm, experiments become continuously learning systems: design and operations are optimized over time, more information is retained and understood, and scientists spend less time on mechanical steps and more on scientific interpretation - directly advancing the national goal of dramatically increasing research output and discovery. This approach also maximizes the return on current investments and enables next-generation facilities to be conceived and operated with AI as a core capability.

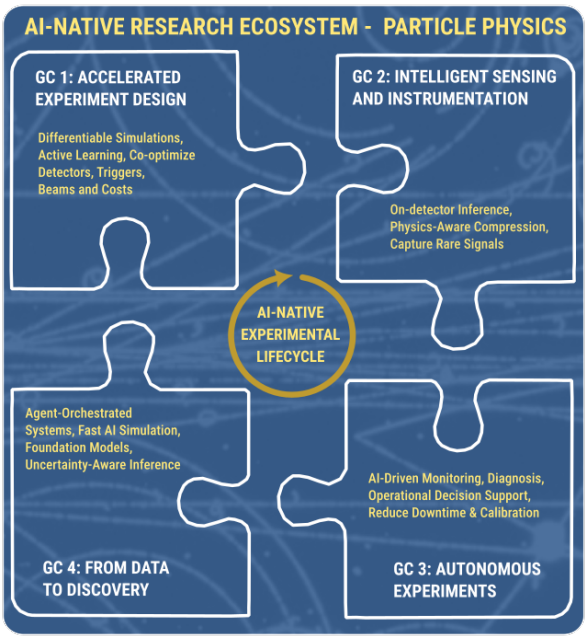


Figure 1: Grand Challenges spanning the experimental lifecycle in particle physics.

Grand Challenges: This vision is organized around four Grand Challenges that form a self-reinforcing engine for an AI-Native ecosystem. *Accelerated Experimental Design* uses differentiable, agent-guided optimization to co-design accelerators, and detectors so that ambitious ideas become buildable, higher impact experiments on faster timescales with reduced technical risk and costs. *Intelligent Sensing & Instrumentation* moves intelligence upstream via trigger-less or AI-assisted readout, physics-aware compression, and real-time inference that preserve rare, unexpected, or time-critical signals while operating within bandwidth and storage constraints. *Autonomous Ex-*

periments transforms labor-intensive, reactive operations into proactive, resilient, and continuously calibrated systems, capturing more high-quality data with less downtime and preserving institutional knowledge over decades-long experiment lifetimes. And *From Data to Discovery* integrates foundation models, fast AI-enabled reconstruction and simulation, and agent-orchestrated workflows to compress analysis cycles by orders of magnitude and open new regions of theory space to exploration. Advancing any one challenge amplifies the others expanding overall scientific reach and productivity while positioning the U.S. as a global leader in AI-powered particle physics.

National-Scale Collaboration: To realize this vision, we propose a *national-scale collaboration* that brings together DOE national laboratories, U.S. universities, and industry partners to develop and deploy shared AI capabilities, impact multiple experiments, and train an AI-literate scientific workforce. Experimental particle physics is unique among scientific fields in its ability to leverage decades of experience building and managing national-scale collaborations. Modeled on successful U.S. operations programs, we envision a national core effort of 120 FTEs or larger, to be jointly managed over multiple large funding partners, and guided by Grand Challenges to build an AI-native research ecosystem. An explicit collaboration avoids duplication while accelerating discovery and technological impact. DOE national laboratories would provide scale, advanced computing resources, and long-term stewardship of complex projects, while universities would drive innovative R&D and workforce development through deep engagement of students and postdoctoral researchers. NSF's mission and community would allow expansion beyond particle physics and provide pathways to engage other domain sciences and translate the impact more broadly in the U.S. research and education ecosystem. This flexible structure also enables targeted investments by private philanthropic partners in innovation, education, and technology translation. Strategic partnerships with industry would ensure rapid access to state-of the-art AI technologies and best practices, enabling efficient technology transfer and co-design. Together, this collaboration would create a cohesive national capability that strengthens connections between labs, universities, and industry and translates U.S. investments in AI and particle physics into sustained scientific leadership and workforce impact.

Workforce Development: The national-scale collaboration we propose would have connections to nearly all U.S. Ph.D.-granting universities and many undergraduate institutions. The DOE Genesis Mission has a goal of training 100,000 scientists over the next decade. Workforce development activities integrated with such a collaboration could contribute 2100 PhDs and roughly 20,000 undergraduates to that goal, with potential for more via neighboring fields such as nuclear physics.

Leveraging National Infrastructure: Our vision is built on an integrated ecosystem in which advanced AI technologies and national-scale cyberinfrastructure jointly support the full experimental lifecycle. Physics-aware AI capabilities such as multimodal foundation models, fast and differentiable simulation with simulation-based inference, and agentic active learning systems coordinate complex scientific workflows. Delivered through inference-as-a-service and embodied in high-fidelity digital twins, these tools link design, operations, and analysis into a closed loop, transforming particle physics facilities into continuously learning systems and dramatically accelerating the path from data to discovery. Realizing this vision requires a new generation of AI-native cyberinfrastructure that extends the community's strengths in distributed computing and data management. It must support large-scale model training, always-on low-latency inference for real-time operations, and AI-accelerated simulation tightly integrated with High Performance Computing facilities. Central to this effort is a federated, AI ready data ecosystem that follows Findable, Accessible, Interoperable, and Reusable (FAIR) principles and enables cross-experiment learning while respecting governance and ownership. Leveraging national assets such as the DOE Leadership Class Facilities, the emerging American Science Cloud, and partnerships with NSF, universities, and industry, this ecosystem provides shared training, inference, workflow orchestration, and interoperable data platforms, amplifying the scientific return of U.S. investments in particle physics and AI.

1 Science Drivers

Particle physics experiments include some of the largest and most complex scientific endeavors. Thousands of physicists, engineers, technicians and students work together to build and operate facilities producing some of the largest datasets in the world in order to unlock answers to foundational questions about the universe. Our experiments already record data samples of 100s of petabytes of data annually, resulting in datasets that are multiple exabytes in size to be analyzed. In addition, the fraction of data that is currently stored and processed can be much less than one percent of the total data generated. Overcoming these limitations due to data transfer and available computing capacity presents a great opportunity for scientific expansion. Experimental particle physics is uniquely positioned to unleash the potential of the AI revolution for science at scales where these methods are transformational.

The Particle Physics community in the U.S. sets its scientific priorities through the Snowmass community process that serves as input to the Particle Physics Project Prioritization Panel (P5). The most recent P5 report [1], from 2023, articulates a compelling scientific vision for U.S. particle physics over the coming decade, focused on a number of foundational questions about the universe. These science drivers span the smallest known scales and the largest cosmic structures, motivating an experimental program of unprecedented scale, complexity, and duration. P5 prioritizes maximizing discovery from major initiatives such as the High Luminosity LHC (HL-LHC), Deep Underground Neutrino Experiment (DUNE) and Proton-Improvement Plan II (PIP II), and Legacy Survey of Space and Time (LSST) at the Rubin Observatory, while advancing critical R&D toward a future Higgs factory, and ultimately a 10 TeV parton center-of-momentum (pCM) collider capable of probing new physics at unprecedented energies.

Deciphering the Quantum Realm: Higgs Physics and New Phenomena A central P5 priority is to use the Higgs boson as a precision tool to probe fundamental physics scales. Direct searches for new particles and interactions and indirect tests via quantum imprints of new phenomena leverage the HL-LHC program and motivate future collider facilities. These measurements demand exquisite control of detector performance, enormous simulated datasets, and sophisticated statistical inference across vast theory spaces. AI can accelerate this program by enabling automated, optimizable analyses; rapid-turnaround simulation and reconstruction; and global searches for subtle deviations from Standard Model predictions. By reducing analysis latency and expanding the accessible parameter space, AI techniques can directly increase the discovery reach of flagship collider experiments and play a central role in the design of future experiments and facilities.

Deciphering the Quantum Realm: Elucidating the Mysteries of Neutrinos Understanding the nature of neutrinos – their masses, ordering, as well as role in matter–antimatter asymmetry, and (astro)physical processes – is a top P5 science driver. Long-baseline experiments such as DUNE and a rich portfolio of complementary measurements from experiments such as NovA, T2K, the Short-Baseline Neutrino Program (SBN), and IceCube are positioned to address these fundamental questions. Neutrino experiments are characterized by low effective signal rates for key physics channels, complex detector responses, and long operational lifetimes. AI-enabled reconstruction, calibration, and data quality monitoring can substantially improve signal efficiency and background rejection, while autonomous operations can increase data taking efficiency and data quality over decades. These gains would translate directly into faster and more precise answers to some of the most fundamental open questions in particle physics.

Pursue quantum imprints of new phenomena (flavor and precision frontier) The P5 report highlights precision measurements of rare processes as a powerful path to uncover new physics beyond the Standard Model. Belle-II and LHCb form the backbone of this program, delivering complementary sensitivity to rare decays, CP violation, and tests of lepton flavor universality in the beauty, charm, and tau-lepton sectors. AI is increasingly essential to this effort, enabling more efficient triggering, improved event reconstruction, background suppression, and global anal-

yses that jointly control statistical and systematic uncertainties across many channels. In the muon sector, the Mu2e experiment will search for coherent muon-to-electron conversion in the field of a nucleus with unprecedented sensitivity, providing a decisive probe of charged-lepton-flavor violation. AI techniques can enhance signal discrimination, suppress backgrounds from cosmic rays and beam-related processes, and optimize operations and calibration over long running periods. Looking ahead, the P5 report emphasizes R&D toward an advanced muon facility. AI will be a critical tool for ultra-low background operation and rare event searches across the muon program. Together, these experiments exemplify how precision measurements, accelerated by AI, has the potential to reveal new fundamental physics through subtle quantum effects.

Illuminating the Hidden Universe: Dark Matter, Cosmic Evolution, and Beyond "Traditional" Astronomy The P5 report emphasizes a diverse experimental strategy to determine the nature of dark matter, to understand what drives cosmic evolution, including dark energy, inflation, and structure formation, and some of the most cataclysmic processes in the universe. This portfolio spans terrestrial direct-detection and axion experiments (e.g., DarkSide-20k, LZ, SuperCDMS, XENONnT, ADMX), accelerator-based probes, and astrophysical/cosmological observations. IceCube anchors the multi-messenger neutrino program and, together with next-generation upgrades (e.g., IceCube-Gen2), enables indirect dark-matter searches and real-time studies of cosmic accelerators and transients. AI is uniquely suited to this overall landscape: it enables anomaly detection beyond predefined signal models, rapid cross correlation of disparate data streams, and real-time responses to transient cosmic events. In this way, AI can expand sensitivity not only to well motivated theories, but also to the unexpected. Dark energy observatories share many common challenges in data volume, complexity, and analysis, as described in a dedicated LSST and Dark Energy Science Collaboration (DESC) AI/ML white paper [2]. This document concentrates on experimental particle physics facilities and science drivers, while recognizing strong opportunities for cross-fertilization of AI tools, infrastructure, and best practices across these domains.

Nuclear Science Drivers: Understanding Visible Matter and Probing Lepton Number Violation The 2023 Long Range Plan for Nuclear Science (LRP) [3] complements the P5 vision by elevating the Electron Ion Collider (EIC) and a multi-isotope neutrinoless double beta decay campaign as priorities to understand how visible matter is built and whether neutrinos are their own antiparticles. The primary science driver for the EIC is to understand how the strong force gives rise to the mass, spin, and internal structure of visible matter. It will map the distributions and dynamics of quarks and gluons inside protons, neutrons, and nuclei with unprecedented precision. The EIC has the potential to become the first major facility to incorporate AI at all stages of design, construction, and operation. The primary experiment at the EIC, ePIC, features a compute-detector integration that enables seamless data processing from detector readout through physics analysis, with the goal of enhancing scientific precision and accelerating the path to discovery. This integration is based on streaming readout and AI driven workflows, and it requires the implementation of AI algorithms to support adaptive, low latency, data driven decision making at a scale and complexity beyond those of traditional approaches. The EIC is naturally aligned with AI-driven reconstruction and analysis because its physics program relies on fully differential, high rate measurements across complex final states.

Both the P5 report and the LRP recognize neutrinoless double-beta decay as an important "quantum imprint" measurement that complements neutrino oscillation experiments and collider searches. It seeks to reveal whether neutrinos are their own antiparticles and whether lepton number is violated; answers with far reaching implications for the origin of mass and the matter–antimatter asymmetry of the universe. Because the signal is extraordinarily rare, backgrounds must be controlled at unprecedented levels. Accelerating this drive for precision discovery in ultra rare measurements would particularly benefit fromt he potential use of AI in analysis and operations.

Enabling the Long-Term Vision: Future Facilities and Sustainability Both the P5

report and a recent National Academy of Sciences [4] report underscore the importance of pursuing an ambitious long-term vision. A Higgs factory such as the FCC-ee will be a crucial step toward fully revealing the secrets of the Higgs boson within the quantum realm and will be a sensitive probe of the quantum imprints of new phenomena. Proposed multi-messenger facilities, like IceCube-Gen2, will probe the furthest and most energetic corners of the universe, while also showing how neutrinos contribute to cosmic mechanisms. On a longer timescale, a 10 TeV pCM collider such as a muon collider potentially sited in the U.S., FCC-hh, or a wakefield-based e^+e^- collider would enable a comprehensive physics portfolio that includes ultimate measurements in the Higgs sector and a broad search program. In each case, AI-driven design and optimization can reduce technical risk, improve cost and schedule realism, and enable holistic co-design of accelerators, detectors, and computing infrastructure. AI can make large experiments more sustainable by reducing labor intensive operations and preserving institutional knowledge, both critical factors for projects spanning multiple decades.

Finally, small-scale experiments play a critical role in our scientific ecosystem and are a natural focus area for AI-enabled innovation. These experiments, often with targeted physics goals and short lifecycles, offer exceptional agility for deploying and validating new AI approaches in reconstruction, simulation, calibration, operations, and analysis. Within the Advancing Science and Technology through Agile Experiments (ASTAE) program highlighted by P5, AI can dramatically accelerate experimental design, reduce labor-intensive operations, and enhance sensitivity in resource-constrained environments. Moreover, successful AI methodologies developed and demonstrated in these projects can be transferred to major facilities, amplifying their impact while minimizing risk.

2 Vision: AI-Native Particle Physics

Our science drivers define what we seek to understand about the universe. Facilities and techniques define how rapidly, broadly, and effectively we can pursue those goals. As experimental particle physics enters an era of unprecedented scale and complexity, AI has become a foundational capability essential to realizing the full scientific potential of our facilities. We envision a future in which AI is embedded end-to-end across the entire experimental lifecycle – from the design and optimization of future facilities, through intelligent sensing and autonomous operations, to simulation, data processing, analysis, and scientific discovery. In this vision, AI is not merely a supporting tool, but a strategic accelerator of scientific impact: managing complexity at unprecedented scales, enabling leaps in productivity, dramatically shortening time to insight, significantly increasing experimental uptime and efficiency, and opening pathways to new transformative scientific capabilities. In short, AI becomes a cornerstone of the future of experimental particle physics and a prerequisite for maintaining U.S. leadership in data-intensive science.

To make this vision actionable, we articulate four Grand Challenges that together cut across current and future particle physics experiments. As summarized in Fig. 2, these challenges are:

- **Grand Challenge 1: Accelerated Experimental Design** - Use of differentiable/surrogate simulations, and active learning to co-optimize detectors, triggers, beams, and costs, informing design choices for future experimental facilities.
- **Grand Challenge 2: Intelligent Sensing and Instrumentation** - Moving intelligence upstream with on-detector inference, trigger-less/AI-assisted readout, and physics-aware compression to capture rare or unexpected signals without overwhelming bandwidth or storage.
- **Grand Challenge 3: Autonomous Experiments** - Automating facility operations and calibration with AI-driven monitoring, diagnosis, and operational decision support to reduce downtime, shorten calibration cycles, and preserve institutional knowledge.
- **Grand Challenge 4: From Data to Discovery** - Building agent-orchestrated, goal-directed analysis systems that integrate foundation models, AI-accelerated reconstruction and simulation, and uncertainty-aware inference to dramatically reduce analysis latency, increase scientific productivity, and expand discovery reach.

Advancing any one Grand Challenge lifts the others (e.g., faster simulation results in better design and search reach), creating a self-reinforcing pipeline that improves sensitivity, reduces latency, and increases return across the science portfolio. Together these Grand Challenges provide a natural organizing principle for a national scale, multi-institutional collaboration that unites universities, laboratories, and industry partners. Shared capabilities, reusable infrastructure and cross-cutting expertise will enable such a collaborative effort to integrate these novel pipelines into scientific best practices used by experiments. The community can then rally around a unifying goal: embedding AI into scientific methodology in facilities across experimental particle physics.

In our vision, the design and operation of experiments is fundamentally transformed by advances in AI. In an "AI-Native" paradigm, facilities, data analysis, simulations, and operations are co-designed with intelligence from the outset, enabling continuously optimized performance, resilient and adaptive operations, and inference that tightly connects data to theory. Near-term demonstrators establish these capabilities at scale; mid-term efforts translate them to standard practice; enabling the next generation of particle physics experiments to be fully AI-native facilities. Demonstrators at flagship facilities can target parts of the experimental lifecycle: HL-LHC and LBNF/DUNE operations; FCC-ee design studies, IceCube-Gen2 deployment, or Muon Collider R&D; while small- and mid-scale agile experiments could be a basis for rapid testbeds covering the entire experimental lifecycle. This staged and deliberate trajectory reflects the community's ambition and aligns with P5 guidance to maximize the scientific return of today's investments while

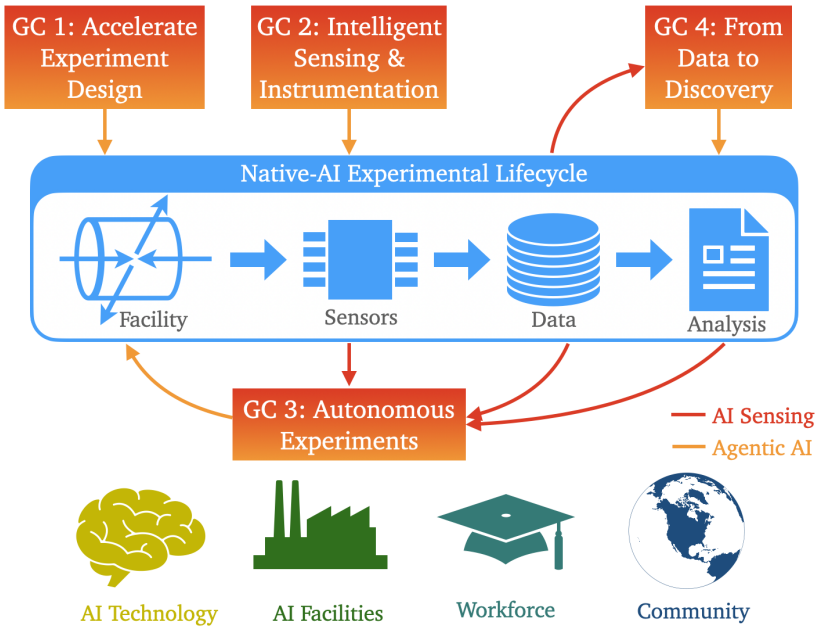


Figure 2: Diagram of the AI-Native experimental lifecycle including the facility/detector, sensing/instrumentation, data acquisition/curation, and analysis. The highlighted Grand Challenges identify where transformative advances enabled by AI accelerate the realization of this AI-Native vision. The lower panel illustrates the shared technologies and infrastructure that underpin and sustain the ecosystem.

building the technological foundation for transformative, next-generation experiments that will define the future of particle physics.

3 Grand Challenges

In this document we are framing our vision around a set of *Grand Challenges*, rather than individual projects. The Grand Challenge approach is better matched to developing long-term road-maps with shared benchmarks and infrastructure, focuses R&D on real-world impact, and encourages both collaboration and healthy competition. It also raises the visibility of the effort to attract talent from within and beyond the field. Most importantly, it creates a compelling narrative to engage the community and potential resource providers.

In preparing the following list of Grand Challenges, we build on several recent community efforts, including the report of the Computational Frontier Topical Group on Machine Learning from the Snowmass 2021 community planning process [5]. More recently, the Accelerated AI Algorithms for Data-Driven Discovery (A3D3) institute hosted a workshop on AI to Accelerate Science and Engineering Discovery in October 2023 and produced a corresponding report [6]. Similar topics were developed by the NSF Institute for Artificial Intelligence and Fundamental Interactions (IAIFI), which produced a report [7] on Generative AI and Discovery in the Physical Sciences. There was also a community report from the NSF Workshop on the Future of Artificial Intelligence and the Mathematical and Physical Sciences (MPS) held in March 2025 [8]. We also note the emergence of curated community resources on the use of machine learning in high-energy physics, such as the High Energy Physics (HEP) ML Living Review [9, 10], as well as focused resources on specific topics including simulation-based inference [11]. As this is a fast-moving area, we also requested inputs from the particle physics community regarding this specific vision document through multiple forums organized by the APS Division of Particles and Fields (APS DPF). The aim of this request for input was to assemble a "big picture" view of the community vision rather than an exhaustive review of all ongoing AI/ML activities and projects. Between December 2025 and January 2026, we received more than 100 contributions from roughly 150 individuals representing 50 institutions (7 national laboratories and 43 universities), outlining opportunities for AI/ML to advance particle physics. The community has significant interest in this area: it is clear that a longer period for input would have resulted in many more contributions from an even greater number of colleagues from across the field. That said, the submissions spanned a broad range, from highly targeted project ideas to more ambitious strategic visions and represent a sampling of the field. Some focused on opportunities or gaps in specific experiments, subfields (such as collider or neutrino physics), physics topics, or detector technologies, while others emphasized opportunities with impact across the field as a whole. The breadth of these contributions and the many efforts in recent years, together with the depth of expertise they reflect, underscore the community's strong potential to leverage AI/ML for transformational advances in experimental particle physics.

3.1 Grand Challenge 1: Accelerate Experimental Design

The main goal of this challenge is to inform design choices for future experimental facilities through the use of differentiable/surrogate simulations and active learning to co-optimize detectors, triggers, beams, and costs.

Particle and nuclear experiment facility design requires years of dedicated expert intuition to tune thousands to millions of often highly interdependent parameters, which can typically only be achieved through slow iterations and greedy component-wise optimization. AI-based design and optimization presents a new avenue to address these critical challenges of scale and complexity and, crucially, navigating this complexity may lead physicists to new and unexpected designs. Future experimental facilities in particular, which are unconstrained by existing hardware and operations, present unique opportunities for AI-powered conceptual design and technology development.

AI-driven optimization enables the simultaneous exploration of many accelerator and detector design spaces, capturing complex interdependencies that are difficult to address with traditional

methods. Detector layout and design can be optimized together with machine parameters, enabling detector performance metrics or even physics analysis capabilities to enter directly into end-to-end optimization loops. This AI-native approach is illustrated in Fig. 3 Technologies such as high-power targetry, high-field magnets based on high-temperature superconductors, and advanced radio frequency acceleration systems can be similarly optimized within a unified and differentiable framework.

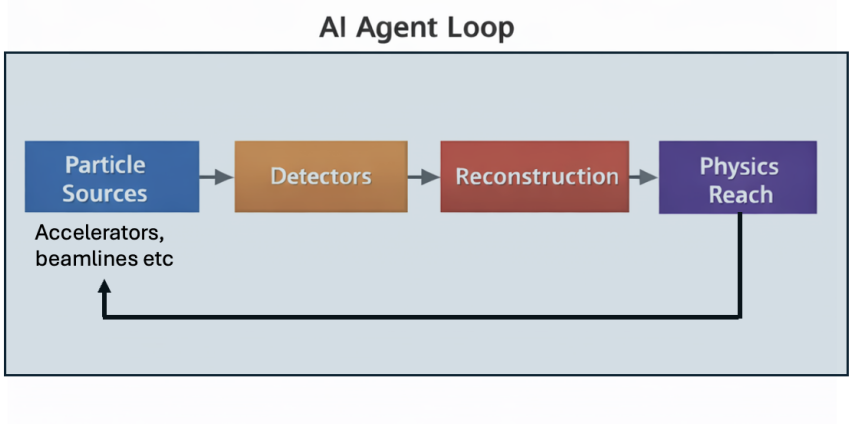


Figure 3: Conceptual schematic of an AI-driven design and optimization loop for future particle physics facilities.

For the most technically mature large-scale future projects, AI-based systems can re-optimize beam configurations and detector layouts to improve experimental sensitivity and technical robustness. As an example, at LBNF/DUNE Phase II, beamline, target and detector configurations involve a high-dimensional space with huge impact on the physics reach of the experiment. At FCC-ee, prototype detector geometry and technology, as well as reconstruction techniques, must still be optimized within the framework of the existing accelerator design to reach the ambitious precision level for achieving the physics goals. AI-driven approaches to these optimizations provide medium-term, high-impact applications of AI that build directly on existing simulation and design workflows.

The R&D path towards a future muon collider, recommended by both the P5 and National Academies EPP reports, could provide the ultimate application of this approach, with the potential to develop the first AI-native experimental facility. In contrast to DUNE and FCC-ee, where several important design elements remain under development but the overall concepts are well established, many key technologies for a muon collider have yet to be fully designed or demonstrated, and the ultimate physics reach and reliability of the facility depends on strong and non-trivial correlation among its many subsystems. Accelerator design, beam dynamics, machine-detector interface, and detector layout are each dictated by millions of interconnected parameters which should be holistically optimized to achieve viable and competitive performance. This approach could similarly benefit other proposed 10 TeV pCM machines, such as FCC-hh or a potential wakefield machine.

Smaller and mid-scale agile experiments with targeted physics programs, including those highlighted in the P5 report under the ASTAE program, provide an abundance of similar opportunities, and often share personnel with the large-scale facilities. These experiments can directly benefit from AI tools, workflows, and best practices developed within larger collaborations, enabling performance optimization, faster design iteration, and improved cost and schedule efficiency at a scale appropriate to their scope. They can also serve as smaller-scale test-beds for new approaches, which can then be adopted by the large collaborations. The transfer of AI enabled methodologies

across experimental programs amplifies their overall impact and helps build a more coherent and sustainable ecosystem for future facilities. An example of such effort is PIONEER [12], a small scale high-rate experiment to study the rarest decays of the charged pion. At the heart of the proposed apparatus, PIONEER features a high-granularity 5D-tracking target, a first for this research program. The use of agentic-AI will significant improve the detector design feedback loop, optimizing the PIONEER's sensitivity to rare decays.

To enable these approaches at scale, AI-based design and optimization tools are critical for both the scale and complexity of the challenges. Technologies such as differentiable simulations and surrogate models can enable gradient-based optimization, while active learning approaches like reinforcement learning and Bayesian optimization can help drive broader exploration of design spaces. To orchestrate such a design loop, agentic AI systems can be used to develop automated workflows that coordinate simulation, optimization, and analysis across multiple subsystems. Traditionally, these challenges have been addressed by disparate communities with disconnected tools, which limits the possibility of holistic optimization. Agentic systems can streamline and simplify complex design and simulation choices, manage iterative optimization cycles, and rapidly evaluate alternative configurations. By reducing manual intervention and enabling adaptive decision making, agentic AI can significantly accelerate the identification of optimal designs in large and complex parameter spaces.

Beyond technical performance, large particle physics facilities operate at the multi-billion-dollar scale, making accurate and reliable cost estimation an essential component of the design process. AI-based large language models provide a powerful opportunity to build robust cost models by integrating empirical scaling laws, detailed engineering estimates, and data from past and ongoing large scale projects. By synthesizing heterogeneous sources of cost and schedule information, such models can support more reliable projections of total cost, construction timelines, and technical risk, and enable rapid evaluation of design trade offs from both performance and cost perspectives.

AI-native experimental design can reduce the technical and fiscal challenges of bringing these projects to life, and introduce new scientific capabilities. Developing common AI-enabled tools for performance optimization, automated workflows, detector design, and cost modeling would dramatically accelerate design iteration, reduce technical risk, and shorten the path from concept to construction.

3.2 Grand Challenge 2: Intelligent Sensing and Instrumentation

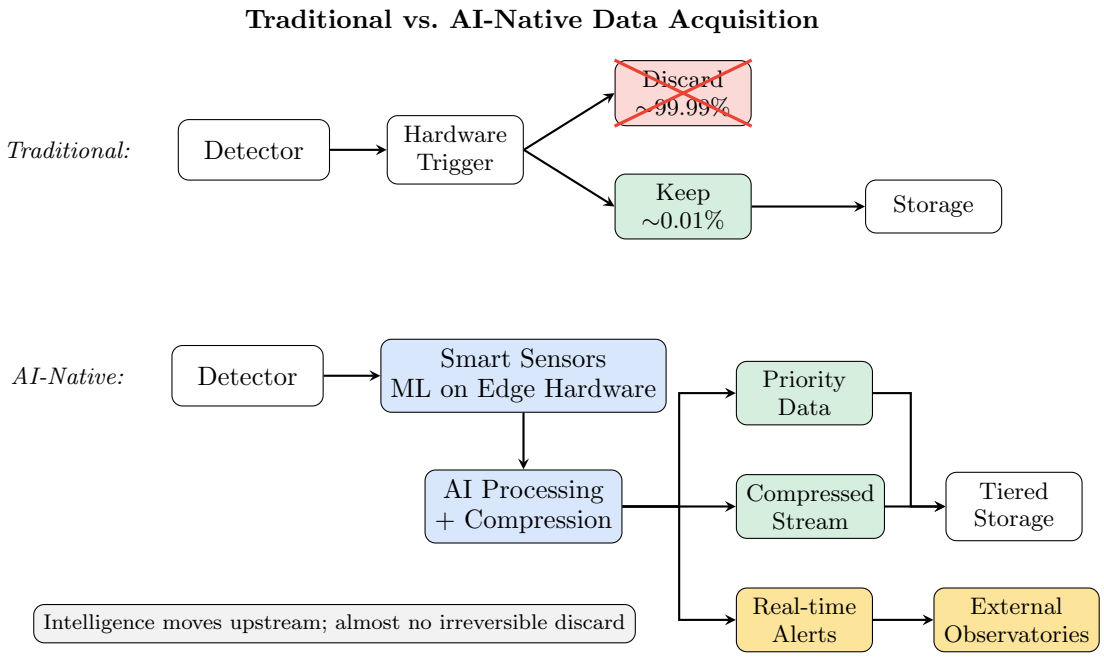
Next-generation particle physics experiments are increasingly limited by how rapidly and intelligently data can be accessed, filtered, interpreted, and acted upon. AI-driven rapid data access is therefore becoming a critical enabler across the full chain of data acquisition, triggering, reconstruction, and analysis, with direct impact on physics sensitivity, discovery potential, and operational efficiency.

A transformative development enabled by AI is the emergence of trigger-less or trigger-free data acquisition architectures, in which detectors operate in continuous readout mode and essentially all data are collected. In this paradigm, traditional hard trigger decisions, in which the majority of data are irreversibly discarded based on simple selection criteria—are replaced by real-time, AI-driven reconstruction and inference. Machine-learning models operating close to the detector can perform fast pattern recognition, timing reconstruction, and physics-aware feature extraction, enabling intelligent prioritization, compression, and routing of data without prematurely rejecting events classified as "background." This fundamentally changes how experiments balance bandwidth, storage, and physics sensitivity, allowing rare, unexpected, or poorly modeled signals to be retained and studied. Realization of this goal will require developing flexible neural networks deployable in on-detector front-end and off-detector electronics, moving intelligence upstream.

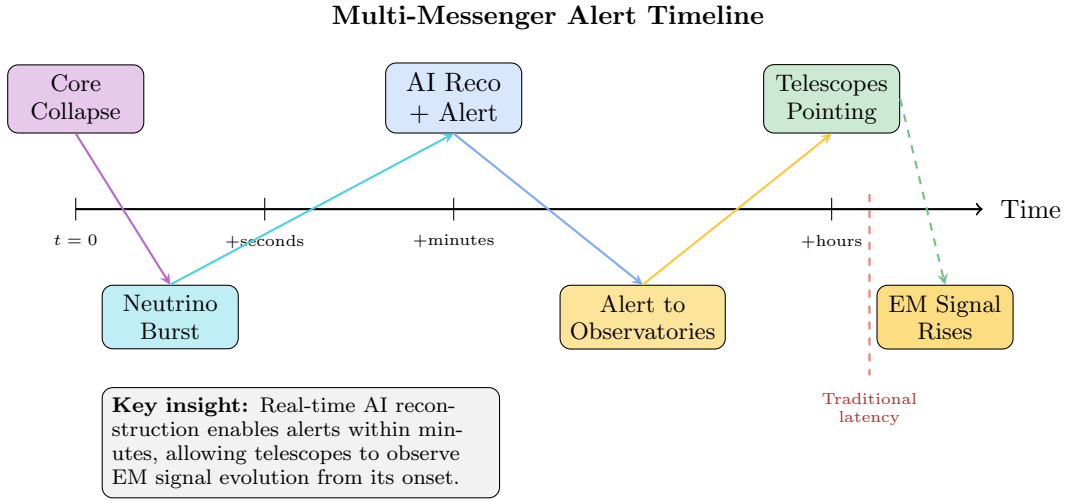
These approaches build on advances in smart sensors and electronics, including smart pixels and ML-assisted front-end systems (ML on Edge Hardware), which embed inference directly into the data acquisition chain. Combined with AI-guided reconstruction and high-dimensional, physics-aware compression, they enable dynamically and intelligently adaptive readout paths that optimize resource usage while preserving scientifically valuable information. Rather than throwing away data at trigger level, AI models learn compact representations that retain subtle correlations and rare features, enabling both targeted physics measurements and anomaly detection beyond predefined signatures.

Trigger-less, AI-driven architectures are particularly impactful for time-critical and data-intensive domains such as fast timing, neutrino physics, and multi-messenger astronomy. In the detection of transient astrophysical events—such as supernova neutrinos—real-time AI reconstruction enables rapid identification and characterization of signals and the immediate dissemination of alerts to external observatories, including optical telescopes, satellites, and gravitational-wave detectors such as LIGO. A rapid localization of the gravitational wave and/or neutrino signatures enables telescopes to be in position to search for the electromagnetic (EM) signature before it even begins to rise (the EM signal is typically delayed by a few hours due to the density of the surrounding material). Thus they can capture the electromagnetic signal's evolution from the very beginning, which contains a rich amount of physics and is often not captured. This capability is essential in constrained environments such as underground (e.g. DUNE) or remote (e.g. IceCube, LSST) experiments, where power, networking, and latency impose strict limits on on-site processing capabilities, data movement, and storage. AI tools could also speed up localization by routing data processing tasks away from sites experiencing computing cluster downtimes or poorly performing network routes. In this arena, every minute counts. Technologies for ML on Edge can be expanded to other domains that use remote sensing, such as geophysical distributed sensor monitoring networks such as SAGE [13] and GAGE [14] and biological and environmental monitoring networks such as NEON [15].

Beyond immediate decision making, AI-enabled continuous readout makes it possible to selectively retain information that is currently impractical to store at scale, such as raw detector waveforms relevant for dark matter searches, solar neutrino measurements, neutrinoless double-beta decay searches, or core-collapse supernovae. By shifting intelligence from rigid trigger logic to adaptive, learning-based systems, experiments gain access to new classes of observables and entirely novel analysis strategies, while simultaneously improving data-taking efficiency, detector uptime,



(a) Comparison of traditional trigger-based data acquisition, which irreversibly discards the vast majority of data at the hardware level, versus AI-native continuous readout with intelligent on-detector processing, compression, and prioritization.



(b) Multi-messenger astrophysics timeline illustrating how real-time AI reconstruction of neutrino events enables rapid alerts to external observatories. Telescopes can be positioned before electromagnetic counterparts begin to rise, capturing physics that would otherwise be missed due to traditional processing latencies.

Figure 4: Intelligent sensing and data acquisition for next-generation HEP experiments.

and long-term scientific return.

Finally, the incorporation of interpretability, uncertainty quantification, and robustness into real-time AI models is a critical emerging direction. As AI systems increasingly determine what data are compressed, prioritized, or permanently stored, transparent and explainable decision making becomes essential for validation, trust, and physics insight. Together, trigger-less architectures and AI-enabled rapid data access represent a paradigm shift in how particle physics experiments operate, enabling faster discovery, greater sensitivity to the unexpected, and more efficient use of large-scale detector and computing infrastructures.

3.3 Grand Challenge 3: Autonomous Experiments

This Grand Challenge targets using AI to improve and automate detector operations, computing operations and calibration derivation that together consume substantial human effort through complex tasks and 24/7 expert support. High-quality, stable data is a prerequisite for physics analysis, and no downstream AI can recover information lost to poor data quality or downtime. Operational excellence underpins all other AI applications: even small gains in uptime, efficiency, or calibration speed translate directly into physics and discovery impact.

Figure 5: Data-taking, data quality and calibration flow diagram and how online and offline information might be used to reduce operational need, expert callbacks and calibration robustness.

This vision aims to enable routine experiment operation with control-room staffing limited to detector safety and decreasing reliance on detector-expert interventions by roughly an order of magnitude. Distributed computing operations, data quality monitoring and calibration would be handled by automated systems, increasing data throughput and reducing the time to derive calibrations from months to days. Together, these advances could deliver up to 50% less downtime, 10× faster calibration cycles, and improved data quality through AI-driven monitoring, diagnosis, and operational decision support—while reducing the required personnel effort for calibration and data quality tasks by a factor of ten.

Today's operations are personpower-intensive and fragile. Large detectors rely on hundreds of experts for 24/7 monitoring and triage. Each hour of detector downtime can cost on the order of $\sim$ \$300k in lost physics opportunity and increases the chance of "missing" the once-a-century/millennium event. Maintaining continuous around-the-clock operations places substantial strain on personnel: shift work drives fatigue, errors, and steep training demands, increasing reliance on scarce expert support. Critical know-how is concentrated in a few individuals, and documentation is fragmented and often outdated. Data quality relies on teams manually reviewing thousands of fast-updating histograms with limited automation, while calibration can require 100+ people and takes months to years of effort, delaying physics analyses. Problems are often discovered only months after data taking, when recovery is impossible, and the effort required to recalibrate limits how often datasets can be improved.

AI will directly reduce downtime and labor by shifting operations from reactive to proactive. Anomaly detection can flag problems before they impact data, while operations assistants based on Large Language Models (LLMs) can combine real-time detector information with logbooks, log files and documentation to guide shifters based on previous events, shorten training, and reduce expert callouts. Predictive maintenance can forecast component failures hours to days ahead, enabling planned interventions instead of emergency downtimes.

AI will also modernize data quality and calibration. Automated Quality Assurance and control (QA/QC) can flag outliers for rapid human review, and real-time physics-level monitoring can provide immediate feedback on data usability rather than delays of hours to weeks. Agentic cali-

bration workflows could shrink week-long campaigns to hours, improving consistency and reducing systematic uncertainties. Similarly, AI-driven assistants and agentic systems can play a transformative role by supporting distributed computing operations, including automated workflow diagnosis, intelligent resource utilization, user support, and real-time monitoring across heterogeneous infrastructures.

These tools also play a critical role in both preserving institutional knowledge as senior experts retire and training future generations of physicists at scale. AI systems can be trained to capture procedures, diagnostics, and lessons learned, providing near-term gains in current experiments while strengthening the foundation for all downstream AI-driven physics. These tools could offer 24/7 interactive access and deep scientific knowledge about facilities.

Key required capabilities include integrated multi-modal time-series analysis across subsystems, reliable root-cause inference for coupled failures, end-to-end automated pipelines, and uncertainty-aware anomaly detection that avoids alert fatigue. LLM assistants must be grounded in detector operations to prevent hallucinations and will require training on internal experiment information.

Near-term progress should focus on deployable pilots: cross-experiment operations AI working groups, subsystem-level automated calibration demonstrators, automation of workflow triage in computing operations and logbook-integrated shifter assistants. By HL-LHC commissioning and first beam at DUNE, experiments could target $\sim$30% downtime reductions, continuous automated calibration, and AI-supported training. The long-term goal is supervised autonomy—experiments designed from day one around AI-assisted operations so humans spend less time responding to problem reports and more time doing physics.

3.4 Grand Challenge 4: From Data to Discovery

This grand challenge seeks to use AI to enable a paradigm-shifting breakthrough in particle physics through innovative data processing and analysis workflows. Today, a single high-energy physics (HEP) analysis typically requires several *years* of sustained human effort dedicated to iterative development, tuning, validation and review. We aim to reduce the time required to design, refine, and evaluate an analysis by factors of 100–1000. These approaches will enable the community to explore a scientific landscape that is itself orders of magnitude larger. Instead of testing a few hand-chosen benchmark scenarios, analyses could efficiently probe entire spaces of theoretical possibilities, mapping the behavior, capabilities, and limitations of whole families of models in a controlled and reproducible way. The proposed methods shift effort away from low-level technical mechanics toward formulating scientific questions, guiding optimization, and interpreting results—fundamentally transforming how HEP conducts science and extending its reach.

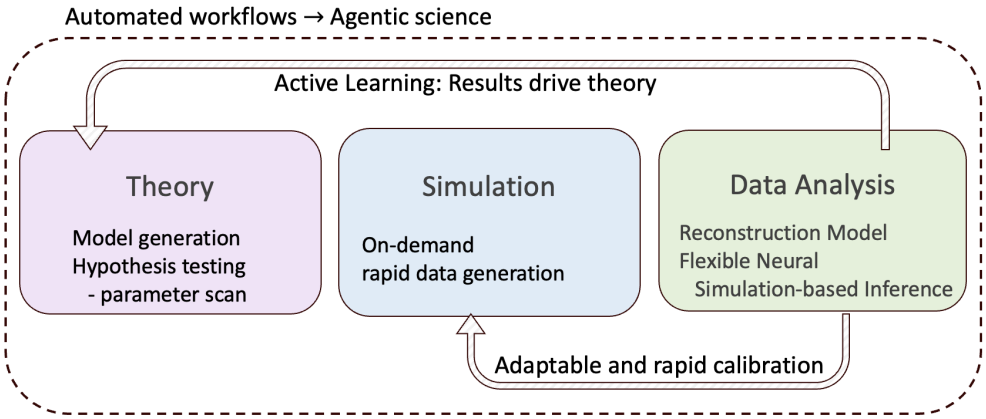


Figure 6: Automatic and optimizable data analysis diagram, adapted from [16]

Realizing this new paradigm requires building systems capable of rapid, optimal, and automated searches and precision measurements across large datasets and extensive theory spaces, following the approach in Figure 6. AI tools can adapt the full analysis pipeline to a given physics objective, optimizing each stage to maximize sensitivity. The **optimization** and **automation** are the two pillars of this grand challenge, and can proceed independently: advances in one improve the performance of the other. Crucially, these techniques apply broadly, not just to one analysis.

Transformational capabilities enabled by multimodal foundation models Multimodal foundation models offer a powerful way to optimize and coordinate the full physics analysis chain, enabling capabilities that are impractical with today's task-specific models. Today's bespoke, narrowly trained models can be replaced with fine-tuning general-purpose multimodal foundation models for analysis-specific objectives. By pretraining on diverse data — including simulation as well as current and archived experimental data — these models can be adapted to new tasks with dramatically reduced labeling requirements and substantially shorter development cycles. A defining feature of modern foundation models is their ability to transfer learned representations across interaction channels, event topologies, detector conditions, and even detector geometries. This transferability is further enhanced by their multimodal nature, allowing joint reasoning over heterogeneous inputs such as images, waveforms, point clouds, and structured physics features.

To fully realize these capabilities, adaptable foundation models must be developed at multiple levels of the analysis stack. At the reconstruction level, they process raw sensor data to enable robust and reusable physics object reconstruction. At the population level, shared representations

improve signal discrimination and background characterization across analyses. At the inference level, foundation models can underpin simulation-based inference, unfolding, parameter estimation, and anomaly detection. This requires dedicated supervised and self-supervised training strategies across all data abstractions of the data—from sensor-level measurements, to reconstructed particles, to full events—while preserving physical interpretability and uncertainty quantification.

AI-Accelerated Simulation for Analysis Simulation is central to nearly all particle physics analyses as an essential bridge between underlying theory, detector response, and observed data. Modern experiments demand simulated samples that are not only accurate, but also rapidly generated, adaptable, and closely calibrated to data. In this emerging paradigm, simulation becomes an on-demand analysis component, where both event generation and the calibration procedures and associated uncertainties can be tailored to specific signatures and updated as understanding evolves.

Today, simulation remains a major bottleneck. Precision measurements and new physics searches require billions of Monte Carlo events, consuming vast computational resources and often limiting statistical reach or the evaluation of systematic uncertainties that depend on modeling assumptions. Calibrating simulation to data—essential for controlling systematics—can itself be slow and labor-intensive. AI methods can transform simulation from a static, pre-computed resource into a dynamic, adaptive element of the analysis loop.

Both algorithmic and hardware advances enable this transformation. Algorithmic techniques such as simulation-based inference, generative models, hybrid physics–ML surrogates, domain translation between simulation and data, differentiable simulation, and systematics-aware fast simulation provide orders-of-magnitude speedups while preserving physical fidelity. These methods support on-demand generation of events and enable rapid exploration of parameter and uncertainty spaces that are impractical with traditional Monte Carlo workflows. In parallel, hardware-aware approaches—including GPU-accelerated physics simulation, GPU-native digital twins, and tight integration with HPC facilities—make high-fidelity simulation scalable to the data volumes and detector complexities of next-generation experiments.

These capabilities are particularly critical for DUNE, the HL-LHC, and future collider facilities, where simulated data needs far exceed the growth of conventional computing resources. Importantly, many of these AI-enabled simulation approaches are differentiable, allowing gradient-based optimization and inference to be integrated directly into the analysis. Within agentic workflows, such "smart" simulations enable active learning across the full analysis chain—from event generation through detector response and calibration—effectively treating simulation and inference as a unified inverse problem. This shift promises dramatic efficiency gains and adaptive, statistically powerful analysis strategies that are highly responsive to emerging physics insights.

AI-Accelerated Reconstruction for Analysis Reconstruction plays an important role in determining the physics reach of experimental data. It transforms raw or simulated signals—hits, waveforms, images, and timing information—into analysis-ready objects such as tracks, clusters, vertices, and particle-identification features, shaping both sensitivity and flexibility. For general-purpose detectors, this process is a primary discovery lever across diverse physics signatures.

Building on the current foundation of task-based models and procedural reconstruction algorithms, large multimodal models offer a natural extension by integrating and coordinating these components within a broader reconstruction ecosystem. Such models can be fine-tuned to jointly reason over heterogeneous sensor inputs and the intermediate representations produced throughout the reconstruction chain, learning shared representations of detector response and particle signatures reusable across multiple downstream tasks. Ultimately, this points toward foundation models trained on extensive, heterogeneous collections of sensor-level data and reconstruction artifacts. By learning general and transferable representations, these models enable homogeneous pipelines that

can be shared across analyses or adapted to specific use cases with limited additional training. In addition, there are problems in reconstruction where procedural reconstruction models still obtain superior performance over task-based models, which require either larger models or more advanced AI techniques to solve.

Beyond incremental gains, AI also enables new paradigms that transcend traditional stage-by-stage pipelines. Probabilistic reconstruction infers posterior distributions of latent physics quantities from raw data that naturally propagate uncertainties and correlations through the analysis chain. Related end-to-end or bypass approaches map sensor-level data directly to higher-level physics quantities or likelihoods to reduce information loss and better align reconstruction with physics objectives. Complementing these ideas, standardized AI embeddings for reconstructed objects and events provide compact, detector-aware representations that can serve as a common interface between reconstruction, simulation, and inference across analyses and experiments. Advances in computational scalability and deployment on accelerators—such as GPUs, FPGAs, and ASICs—will further extend these capabilities into low-latency trigger systems and streaming analyses. Integration of agent-orchestrated workflows with fast AI-based simulation and AI-accelerated reconstruction will increase physics reach, flexibility, and discovery potential.

AI for Statistical Interpretation The final stage of an analysis is statistical interpretation, where experimental observations are compared with theoretical predictions. This step is often limited by the difficulty of solving the inverse problem: realistic likelihoods are frequently intractable, forcing traditional approaches to rely on low-dimensional summary statistics that lose significant information. Neural simulation-based inference (NSBI) addresses this by learning the mapping between observables and model parameters directly from simulation, constructing neural likelihoods, likelihood ratios, or posteriors that enable flexible, information-rich inference for both discovery and precision measurements. Related advances are also transforming unfolding, which connects detector-level observations to theory-level quantities at the distribution level, into a modern, more robust and scalable inference task driven by generative models and simulation-based learning. In parallel, anomaly-detection offer a complementary, model-agnostic approach by operating directly on data without relying on specific simulations or theoretical hypotheses.

Realizing the full potential of these methods requires active learning and closed-loop workflows. By using intermediate analysis results to guide theory selection, simulation, and further inference, active learning—together with AI agents—enables scalable automation across complex hypothesis spaces and closes the loop between data, theory, and interpretation.

These methods will also strengthen the interface between theory and experiment, enabling rapid exploration of high-dimensional theory spaces, including effective field theories and simplified models, yielding dynamic interpretations that evolve as new data or theoretical ideas emerge. These approaches can identify poorly constrained regions of parameter space, expose tensions between datasets, and suggest targeted re-simulation or new measurements. Extending these capabilities across experiments enables global interpretations of shared theoretical models and a far broader exploration of particle physics than is feasible today

Agentic workflows Agentic AI automates the scientific workflow itself. Given a high-level goal—such as searching for dark matter or measuring the Higgs self-coupling—a multimodal, agent-based system built on large language models can organize the end-to-end analysis: selecting parameter points to evaluate, incorporating active learning to guide hypothesis exploration, and iteratively refining the analysis strategy as results accumulate. The agent orchestrates the technical tasks, including Monte Carlo generation, analysis configuration and execution, systematic evaluations, and statistical modeling. The physicist directs the scientific process by defining the analysis goals, assumptions, and constraints that guide the automated system. This allows advances in underlying capabilities—such as faster simulation, improved foundation models, or more powerful inference

techniques—to appear immediately in the workflow without manual re-engineering by the analyst. In addition, the use of large language models may offer a complementary interpretability layer that integrates intrinsic, model-level mechanisms with prompt-driven, human-in-the-loop interaction.

Summary: The result is an analysis chain that is not merely automated, but adaptive, interrogable through physicist prompts, optimizable at every stage, and scalable across the full breadth of particle physics measurements and theory space.

3.5 Cross-cutting Themes and Emerging Opportunities

Cross-cutting themes and areas of active community interest: While many of the community's contributions aligned naturally within the four Grand Challenges described above, some of the proposed projects either fall outside those categories and/or have recurring themes that cut across them. These merit explicit acknowledgment and are discussed here. First, neutrino interaction modeling and cross section inference emerged repeatedly, with proposals to combine simulation-based inference, generative surrogates, and domain translation to learn robust, data driven models; and to build public neutrino benchmarks and Open Data challenges that standardize tasks, metrics, and leaderboards. In parallel, inputs emphasized theory–experiment integration, notably Standard Model Effective Field Theory (SMEFT/EFT) inference pipelines and workflows aware of parton distribution functions (PDFs) that connect lattice Quantum Chromodynamics (QCD) and phenomenology to experiment via reusable latent representations and uncertainty-aware inference. A third cluster centers on Monte Carlo including GPU-accelerated event generation. Complementing these are proposals to make data truly AI-ready, with curated, federated Open Data, common tokenized representations, and inference-as-a-service deployments that carry models across facilities; and to embed AI in operations and QA/QC through anomaly-aware data quality monitoring, agentic shift assistants, and vision based fabrication/assembly checks. Finally, several inputs highlight co-design of detectors and materials – from smart pixels and embedded FPGAs to ML-guided discovery of scintillators and optical interfaces – linking instrumentation innovation directly to AI-driven design loops. Together, these threads reinforce the need for solutions that span design, operations, analysis, and long-term stewardship of data and knowledge. The cross-cutting AI technologies and the cyberinfrastructure that underpin these projects and the four Grand Challenges are discussed further in Sections 6 and 7, respectively.

Interplay with theory and the case for coordination: While our whitepaper concentrates on experimental particle physics, our community input underscores that AI-enabled discovery depends on a tight feedback loop between theorists and experimentalists: EFT and PDF advances inform generator developments and analysis objectives; differentiable and generative simulators shorten the path from theory hypothesis to experimental test; and curated, AI-ready Open Data lowers barriers for joint inference across datasets and facilities. The same inputs point to shared needs that exceed any single project: common data/representation standards and benchmarks; portable training and inference services; agentic workflow orchestration; and operations toolkits that reduce downtime and preserve institutional knowledge. These cross-cutting opportunities directly motivate the next section: a larger, coordinated collaboration that federates expertise across labs and universities, pools computing and data services, aligns with emerging national infrastructure, and scales workforce development so that each experiment benefits from shared tools, sustained support, and the ability to marshal greater collective effort on every project.

4 Collaboration - Building a National Effort

We envision creating a single **national scale collaboration** that brings together U.S. national laboratories, universities, and diverse particle physics experiments to build a shared AI-native research ecosystem for particle physics together with industry partners. A national scale collaboration is achievable. The U.S. particle physics community already has extensive experience with large-scale collaborations between U.S. universities and DOE national labs as well as international partners, as shown in Figure 7. For example, the U.S. CMS collaboration includes two DOE labs (FNAL, LLNL) and 51 universities, while the U.S. ATLAS collaboration involves five DOE labs (BNL, ANL, LBNL, LLNL, SLAC) and 42 universities. Over the past three decades, these collaborations, with international partners, have successfully built, operated, upgraded, and delivered scientific results from the LHC experiments, and the DUNE-US organization is largely modeled after them. This national scale collaborative model enables significant resources to be targeted at R&D that moves quickly and strategically, guided by the Grand Challenges, while also providing a fast and structured path for universities to get involved.

In the U.S. experimental particle physics ecosystem, national laboratories and universities play complementary roles. National laboratories provide large-scale production capabilities, access to advanced technological facilities, and highly skilled technical staff necessary for the realization of complex projects. Universities contribute in areas such as cutting-edge R&D, workforce development, and the active involvement of students and postdoctoral researchers, enabling rapid innovation and sustained intellectual vitality. National laboratories can further serve as hubs that foster collaboration among geographically proximate universities, strengthening regional and national research networks. The greatest scientific impact is achieved when laboratories and universities are equal partners and tightly integrated via a common scientific goal to bring together academic innovation and training with laboratory-scale implementation and long-term stewardship.

How would a national scale collaboration work? The success of this collaboration requires more than the typical institution-by-institution "bring your own budget" model used by many traditional scientific collaborations. While bottom-up participation will remain essential, it is not sufficient on its own to support the scale, coherence and sustained effort required here. Thus **joint management** of a larger core effort across multiple large funding sources—DOE (Genesis Mission), NSF, and private foundations—should be pursued. This structure would be modeled on the U.S. LHC operations programs, each of which manages O($40M/year) from two funding partners (DOE and NSF). Though the specific scope for each funding source is well defined, unified management of major U.S. funding streams supports the definition, pursuit, and evolution of coordinated "Grand Challenge" activities, developed in close partnership with the experiments. Leadership is drawn from both labs and universities. An annual planning and budget process (spanning the DOE and NSF streams) updates funding for collaborating institutions based on progress towards milestones and deliverables guided by the Grand Challenges. In the case of the national scale AI collaboration, an additional challenge is delivering to multiple experiments with different scientific objectives and timelines. This will be informed by, and built on, the (cross-experiment) experience from other community national-scale R&D projects such as the DOE-funded HEP-CCE [17] project, NSF-funded Institute for Research and Innovation in Software for High Energy Physics (IRIS-HEP) [18–20] and A3D3 [21, 22] institutes. Experience with the creation of national or regional physical hubs to support distributed collaboration will be drawn from the Fermilab LHC Physics Center, the U.S. ATLAS Analysis Center (at LBNL, BNL and ANL) and the NSF-funded IAIFI AI Institute [23]. Leadership from all of these efforts have been involved in this whitepaper.

What scale of collaboration is needed? Achieving the most ambitious goals for a national AI collaboration in particle physics will require substantial scale. Using the past efforts and structures

of the full U.S. high-energy physics program as a guide, we estimate that a vibrant national program would likely require a core effort (analogous to the operations programs) of at least 120 full-time equivalents over 5 years and involve twice that number of contributors. A core effort of this size or larger, involving both universities and DOE labs, aligns with the scale of existing U.S. collaborations and projects. It is intended as an estimated scale, given the scope and ambition of the Grand Challenges, rather than a detailed bottom-up effort calculation. While the effort will include research components, these activities will be explicitly guided by the Grand Challenges with the aim of building an AI-native research ecosystem supporting the full lifecycle of particle-physics experiments. At this scale, the collaboration core would also serve as an intellectual hub for many smaller separately funded R&D efforts from across the broader community as well as providing a clear point of contact for international and industry partnerships. Such an AI collaboration is not meant to replace, but rather to augment, the existing experimental operations programs.

Collaboration Summary: The experimental particle physics community has built, over many decades, organizational structures connecting essentially every major U.S. research university as well as many undergraduate institutions and nearly all of the DOE Office of Science laboratories. The HEP community's historical head start in data-intensive science, combined with these mature structures and diverse range of experimentation, uniquely positions it to create a national-scale effort. The collaboration we describe will leverage these strengths to develop new methodologies for data- and AI-enabled discovery, while contributing broadly to the physical sciences and to the creation of an AI-enabled workforce.

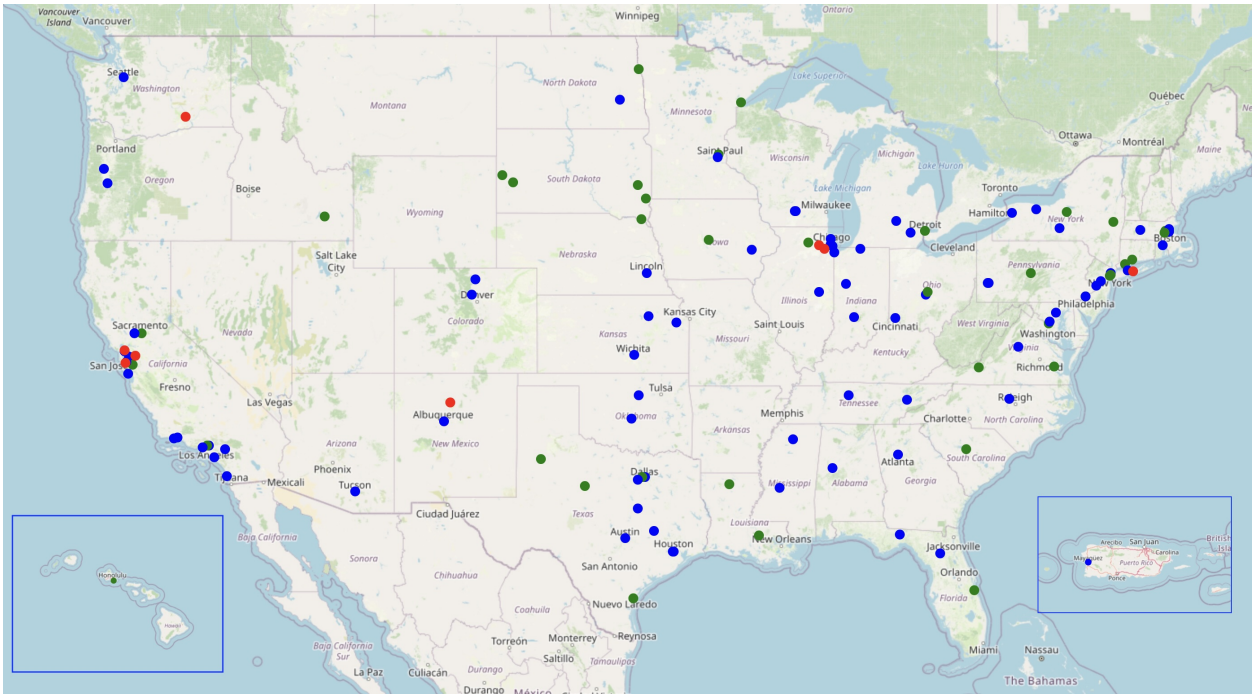


Figure 7: U.S. HEP Institutions (blue dots = universities, red dots = DOE labs) that are currently part of large national-scale funded projects: U.S. ATLAS Operations (DOE/NSF, 33 institutions), U.S. CMS Operations (DOE/NSF, 44 institutions), DUNE Operations (DOE, 16 institutions), HEP-CCE (DOE, 4 labs), IRIS-HEP (NSF, 16 universities), A3D3 (NSF, 12 universities). Additional universities that are part of the U.S. ATLAS, U.S. CMS and (U.S.) DUNE scientific collaborations are shown in green. Taken together, these overlapping experiments and projects tightly connect 8 DOE labs and 120 different universities and colleges.

5 Workforce Development

Training and workforce development is key for the success of this initiative. The national-scale collaboration we propose will be capable of providing a structured pipeline for students and early-career researchers at all levels, from undergraduate through post-doctoral stages. Unifying efforts across universities and labs via such a collaboration will also transform fragmented training into a centralized and coherent ecosystem that both meets the needs of the community and also contributes to building a national AI-enabled workforce. By scaling beyond individual university or lab-level efforts, a national workforce program institutionalizes training and skill transfer, ensuring the long-term sustainability and impact of the AI-enabled research ecosystem.

Partnerships between universities, DOE national laboratories, and industry will be the created around scientific discoveries and technology development. On one hand, the engagement between universities and industry partners will allow universities to serve as think tanks for U.S. technology companies, ensuring leadership in the global market. In the other direction, a close relationship with industry partners will allow students and early career researchers to learn with the most modern AI tools and technologies. Universities and national laboratories have a long tradition of close engagement in workforce development, particularly in particle physics. These ties will be further strengthened by involving students in advanced AI projects in the labs, ensuring that students graduate having had first-hand experience with the development of new technologies.

Particle physics is well suited for dual-competency training with several applications that serve as testbeds for learning and developing AI methods. A key feature of this area is the availability of large datasets which can be used for AI. Large-scale experiments in particle physics make the data available in efficient data formats distributed in several DOE laboratories and universities across the country. These datasets offer a controlled environment to learn and develop new ideas in AI/ML.

Workforce development will naturally proceed in two overlapping and synergistic approaches. Participants will contribute to R&D projects to build the AI-enabled research ecosystem, gaining hands-on experience with foundational AI tools and our national infrastructure. As the projects mature, they will increasingly focus on using these capabilities to accelerate scientific discoveries and technology development. DOE laboratories and Universities should be equal partners in this training pipeline with each managing different aspects in a coordinated fashion. We envision impact in different ways at different levels of education and training.

Undergraduate education: Undergraduates will develop core knowledge and interest in our domain science (particle physics measurement and discovery) and AI-literacy simultaneously. There are several courses on AI fundamentals already available that include topics in mathematics, statistics, and computer science. These courses could be offered as part of the preparation for physics students, supplementing the mathematical methods and computational physics curriculum. The foundational training acquired through coursework is supplemented by particle physics domain-specific training and practical experience that connects the more general courses to specific research and technology development programs in AI for physics applications. Given the breath of this training, we encourage physics students to pursue dual-competency degrees. Traditional undergraduate research programs such as NSF Research Experiences for Undergraduates (REU) and DOE Science Undergraduate Laboratory Internships (SULI), as well as HEP-specific programs such as the Program for Undergraduate Research Summer Experience (PURSUE) and Summer Undergraduate Program for Exceptional Researchers (SUPER) organized by the U.S. CMS and U.S. ATLAS operations programs, and the IRIS-HEP Fellows program, can be enhanced to bridge from academic coursework to a cutting-edge research environment (in both physics and AI tools) and provide a broader context and numerous AI-related projects beyond traditional detector development or data analysis. Undergrads will connect via summer programs and (funded or unfunded) via independent

study and/or senior thesis activities. Post-baccalaureate and bridge programs help recent bachelor degree recipients to gain the research skills and academic credentials needed to become competitive applicants to graduate programs and industry. Opportunities such as the A3D3 post-baccalaureate program [24] can be extended to a broader set of research programs and incorporate both industry and lab partners.

Graduate education (Masters/PhD): Graduate researchers will grow into scientists with dual physics/AI competencies through a formalized co-mentorship model that brings together university faculty, national laboratory staff, and strategic industry partners. Building on and scaling the successful two-year DOE Computational HEP Traineeship programs [25–27], this effort will embed PhD students within large, multi-institutional teams working on the Grand Challenge problems. In addition, we envision targeted annual training modules—often difficult for individual universities or laboratories to offer—that will provide up-to-date, practical instruction for integrating AI across the research ecosystem. The DOE Science Graduate Student Research (SCGSR) program can be expanded to support graduate students to spend an extended period at the DOE national labs to pursue their research program and develop advanced skills. Programs supporting practicums at DOE national labs, such as the DOE Computational Science Graduate Fellowship, can also be expanded to include such practicums in industry partners.

Postdoctoral education: By leveraging their competencies in both particle physics and AI technologies, postdocs will be well positioned to play key roles as technical architects of elements of the research ecosystem. Build leadership skills, prepare for roles as future principal investigators or as senior technical leads in national labs or in industry.

Potential Training Scale: As part of the Genesis Mission, for example, the DOE has described a national goal of training 100,000 scientists and engineers over the next decade "to lead the world in AI-powered science, innovation, and applications", and is currently seeking input on workforce development to achieve this goal [28]. The national-scale collaborative project we describe—organized around Grand Challenges and spanning multiple DOE laboratories and universities across the U.S.—offers a unique and concrete opportunity to contribute significantly to this goal of developing the next generation of AI-literate scientists. The United States has produced an average of 215 PhDs in particle physics each year [29, 30] over the past 15 years, with some annual variation as new experiments begin taking data. The number of undergraduate students potentially engaged in this research is likely an order of magnitude larger. Over the next decade, this equates to roughly 2,000 PhD and 20,000 undergraduate students that would benefit directly from this collaboration and its AI-enabled research ecosystem. If closely related nuclear physics experiments and students were also engaged in the collaboration, these numbers would be 50% higher. This represents a potential significant contribution to the national Genesis Mission objectives. By involving students and early-career researchers directly in the creation and use of such a cutting-edge AI-native research ecosystem, we will simultaneously drive transformative discovery and help secure long-term U.S. leadership in AI-enabled fundamental science.

Building on existing efforts: In addition to the existing workforce development programs mentioned above, the particle physics community has been a leader in building national level programs due to its national scale scientific collaborations and large projects. For example, the DUNE collaboration's workforce development program [31] has one thrust geared towards undergraduate students and another towards early-career researchers. A number of annual summer schools focus on AI and related technology and help build a cohort experience for participating PhD students and postdoctoral researchers. These include the annual IAIFI summer school [32] and the Computational and Data Science Training for High Energy Physics (CoDaS-HEP) summer school [33].

Complementing these efforts, the long running QuarkNet program [34] provides a nationally coordinated education and outreach program that engages high school teachers and students, connects classrooms to frontier particle physics research, and thus further strengthens the long-term STEM workforce pipeline.

6 AI Technologies for High Energy Physics

To drive progress on these grand challenges, several cross-cutting technologies need to be developed or adapted for the particle physics setting. In many cases, progress requires developing new approaches, such as innovative training methods, multi-modality, scaling and adaptability. These AI technologies include:

Foundation Models that are models pretrained in a way that they can be adaptable to a variety of downstream tasks. These models are typically large and trained on massive datasets, necessitating the need to understand *scaling* and the compute requirements for training HEP models. Ability to accommodate multiple data modalities (e.g. different detector components, human language interface) is crucial and effective ways to fuse information across data modalities to extract general features reusable for high precision, complex HEP tasks need further R&D. Several such models will be developed, e.g. for adaptable reconstruction, population discrimination, and physics modeling of detector response and particle tracking that are needed across the grand challenges. Inference on such models is also a challenge, especially when aiming to adapt toward low latency systems like triggers, likely necessitating inference as a service systems. Developing the architecture and training procedures for foundation models requires R&D for physics-native data types, research beyond what has been done in industries, as they come at an unprecedented scale (e.g. millions events, millions sensors per event, 3D image with billions of voxels) for high precision physics inference. Finally, much of the present research on foundation models for HEP focus within each science domain. One large foundation model (Figure 8) that covers all frontiers of HEP research with a common physics background knowledge will require future R&D.

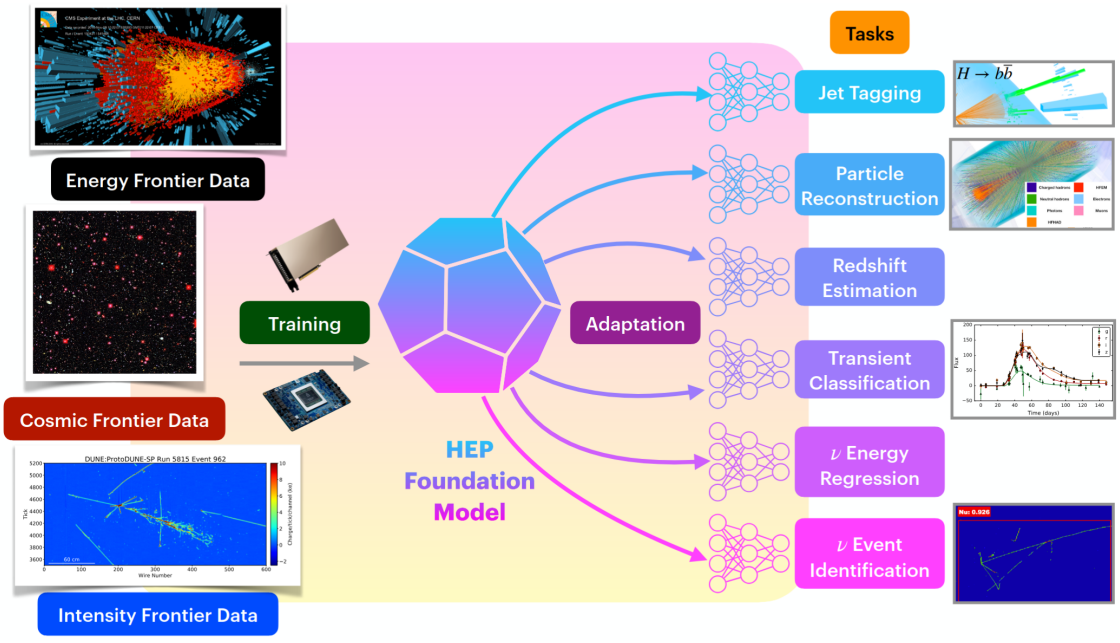


Figure 8: A cross-domain, multimodal foundation model for HEP.

Fast Data Generation / Simulation to provide rapid, adaptable, and ideally on demand data creation for widespread use in HEP analysis such as training and adapting models, statistical inference, and two-sample testing for data quality. These tools are crucial for design optimization, operations, and analysis (grand challenge 1, 3, and 4). Building these tools will likely require

generative AI techniques such as diffusion models or flow-based methods that are adapted to physics data structures, as well as standard simulations that are accelerated through large scale compute systems and workflow automation. Ensuring the fidelity of these new tools will require robust validation procedures using a large set of curated measurements from HEP experiments.

Differentiable Simulation corresponds to integration of simulators into automatic differentiation frameworks. By making such codes differentiable, they become amenable to gradient-based optimization and capable of solving a spectrum of inverse problems. Applications concerning grand challenges 1, 3 and 4 include sensitivity analysis, automated detector calibration and reconstruction, design optimization, and simulation-based inference. Via chain rules, differentiable simulators can become a greater differentiable pipeline (e.g. reconstruction) and effectively enabling an end-to-end optimization.

Neural simulation-based inference includes a variety of methods at the conjunction of simulation, probabilistic modeling, and deep learning and generative AI. These methods enable high-dimensional Bayesian and Frequentist statistical inference that can be signifcantly better than prior methods (e.g. histogram based solutions).

Active learning and reinforcement learning provide paradigms to enable iterative and measurement informed decision making, e.g. for choosing the next hypothesis to test, choosing the next data quality check to perform, or choosing the next step in a design process. These technologies have only seen limited use in HEP, but will play key roles in the feedback and decision making process envisioned here including design optimization (grand challenge 1), facility operations (grand challenge 3) as well as hyper parameter optimization in training AI models.

Agentic systems benefit from both the strong reasoning capabilities of modern LLMs, the strong coding abilities of modern LLMs, and the LLM's ability to run external tools and interpret such codes output. Using such tools "off-the-shelf" or exploring potential fine-tuning (e.g. via reinforcement learning) of such systems for HEP data can provide the critical outer-loop orchestration of workflows and design planning that can drive more automated decision making across all of the grand challenges.

AI Training and Inference as a Service is the backbone for any AI-native and -driven experimentation and data analysis. Without ready access and efficient usage of of AI-focused computing resources any progress enabled by AI methods would be hindered. Integrating a new workflow paradigm, i.e. offloading AI compute to an external service or service provider, is essential to reap the highest benefit from AI-native methods.

On-edge AI focuses on the development and deployment of AI models on platforms such as FPGAs and ASICs where resources are constrained. These technologies seek to provide inference capabilities for increasingly complex models in systems where the latency and/or resources are highly constrained, such as the trigger at colliders or in systems with remote sensing capabilities. Technology developments focus on areas such as compression to reduce the resource footprint and faster inference, while maintaining physics performance.

Digital Twins are models of physical experiments that are realistic, dynamic and that can be queried, adapted or optimized as if they were the actual experiment. These models are the enabling substrate for AI-native experiments, making it possible to turn the experiment into a learning system.

7 Advanced AI Cyberinfrastructure

The resources required to support the **new vision of AI integration** laid out in these grand challenges will be significant. This includes a massive increase in access to **"AI-ready" data**, as well as support for data management and curation, together with **large-scale compute** systems capable of performing AI training on HEP datasets, serving foundation models for inference, LLM hosting for agent-driven experiment workflows including detector calibration, simulation, reconstruction, and assisted analysis. These end-to-end science capabilities must be supported with sufficient resources to accelerate discovery.

Particle physics brings **unique opportunities and skills to the national cyberinfrastructure** (CI). The experiment **exascale data** is organized, managed, transferred, accessed and accounted for using distributed management systems that are able to integrate scientific and commercial storage systems. In addition, the tradition in HEP of large experiments distributed across many institutions has resulted in **decades of expertise in distributed computing**; the computing approach has reflected the organizational infrastructure, resulting in valuing portability between sites, common APIs/services, and workloads that can be partitioned effectively across sites. The HEP community is unparalleled in its ability to be an CI "omnivore", leveraging almost any conceivable computing infrastructure. **HEP's scale in data volume and data flow complexity is unparalleled**: while the large LHC experiments move over an exabyte a year between dozens of computing centers across international boundaries, even smaller experiments leverage the same common services to cross boundaries. Given the aggregation of researchers' workloads by experiment, modest investments to have experiments adopt new services can impact thousands of scientists and serve as a premier training ground for a new, AI-enabled workforce. **Partnerships with industry** are critical to enable **access to state-of-the-art AI tools and technologies**: the HEP community has extensive experience working with industry, particularly through subscriptions for resources and services that can be integrated in the HEP cyberinfrastructure.

These historical strengths mean the community (a) can quickly leverage new resources, (b) is skilled at adopting new interfaces, and (c) ensures that modest investments have amplified impact on science. These attributes make the **HEP community the ideal launchpad for a coordinated national AI infrastructure**. Just like the HEP community collaborates across funding agency, we envision leveraging DOE assets like the Leadership Computing Facilities (LCF) and the American Science Cloud (AmSC) at the largest scale and NSF's breadth of infrastructure across the US university ecosystem. Furthermore, engagement with industry is essential, to create hybrid models that integrate commercial best practices.

Realizing this vision requires **complementing the current paradigm** of "distributed batch processing on massive datasets" **with new services** that allow massive, bursty training of foundation models at the largest of scales, training of large ensembles of medium-sized models for experimentation with new AI methods (optimal for distributing over all available resources), and the always-on, low-latency inference required by agentic or high-throughput inference workflows. Combined, these represent a challenge that, if successful, **leverage and amplify the value the existing federal investments** and is a key enabler for the grand challenges.

Data Curation, Governance, and Open Access Data is the fuel for foundation models and AI at scale, but reconstructed physics data (interaction events) is rarely "AI-ready". To enable cross-experiment AI ingestion and empower agents to easily understand and perform data analysis in support of exploratory interactions driven by scientists, the infrastructure must support a dedicated **Data Curation** layer that transforms these events into coherent, community-defined formats optimized for modern machine learning pipelines.

- **Federated Data Lakehouse and Open Data:** We envision a federated "Data Lakehouse"

that enables data-level queries to extract specific variables or slices of datasets across distributed facilities. This architecture must support **Open Data** principles and provide secure access to high-value community datasets while respecting data sovereignty and ownership policies.

- **Coherent Representation and Formats:** To enable cross-experiment foundation models, data must be curated into coherent, self-describing representations (e.g., tokenized sequences, vector embeddings) with standardized interfaces, ensuring a model trained on DUNE data can be technically interoperable with analysis tools developed for the LHC. The approaches to "tokenize" event data are rapidly evolving and this area would benefit immensely from US leadership.
- **Automated Curation Pipelines:** The infrastructure must provide standardized, containerized workflows to ingest, clean, and annotate interaction events. This includes automated metadata extraction to ensure all data is FAIR (Findable, Accessible, Interoperable, and Reusable) and discoverable by AI agents.

Any one of these items is challenging – however, given the exabyte scale and variety of HEP datasets (a single experiment may have tens of thousands of datasets) and complexity (an event can result in thousands of derived quantities), the opportunity for impact is unprecedented.

Computational Resources and Hardware Adaptation The infrastructure must support distinct modalities of AI computation, leverage the investment in existing physics codes to produce new datasets, and use emerging hardware architectures:

- **Training and Fine-Tuning at Scale:** Developing Physics Foundation Models (Challenge 3.4) requires access to leadership-class computing facilities (such as NSF's LCCF, expected to begin operations in 2026) capable of massive parallel training. This infrastructure must support flexible execution environments for training complex multi-modal architectures and automated hyper-parameter tuning as a service.
- **Increasing Science Throughput:** The HEP community has a breadth of models to train, optimize, and use for large-scale processing. It is uniquely positioned to use all available resources, including mid-scale and university-level investments.
- **Adapting Simulation for AI-Centric Hardware:** While traditional HEP simulations heavily rely on double-precision (FP64) arithmetic, modern AI-accelerated hardware increasingly prioritizes mixed- and low-precision capabilities (e.g., FP16, FP8). To maximize scientific throughput on next-generation supercomputers, we must invest in refactoring simulation codes to leverage these lower-precision units and allow for AI-enhanced or AI-supplemented simulation codes without compromising physical accuracy.
- **Agent-Driven Simulation and Orchestration:** Enabling "Agentic Science" (Challenge 3.3) requires infrastructure that allows AI agents to dynamically provision simulation resources (e.g., event generators, detector simulations) in response to active learning loops. This requires a "Workflow as a Service" capability where simulation containers can be spun up on-demand to test hypotheses generated by an AI agent.
- **Long-Running Inference Services:** Deploying AI agents, foundation models, and AI-native simulation and data analysis codes requires persistent, low-latency inference endpoints. The infrastructure must provide stable hosting for open-source foundation models (e.g., Llama, Mistral) while simultaneously offering secure, integrated access to leading proprietary industry models. This enables researchers to query the state-of-the-art—whether open or commercial—instantaneously within their analysis loops.
- **Integration of edge with distributed resources:** Real-time AI processing enables autonomous experiments, but also needs to be complemented by access to distributed data and

computational resources at scale for tasks such as AI training, inference, and for optimization and validation using simulation or a digital twin.

Leveraging the American Science Cloud (AmSC) for the Genesis Mission The developing **American Science Cloud (AmSC)** offers a unique opportunity to build these capabilities at a national scale. If the High Energy Physics community actively engages with AmSC during its formation, the resulting infrastructure and tools would serve many of the specialized purposes we require, effectively providing the platform for the **Genesis Mission**.

- **Unified Inference for Open Source and Industry Models:** AmSC plans to provide the "Inference as a Service" layer described above, managing the complexity of serving long-running open-source models while handling the authentication and subscriptions required for industry model access.
- **Genesis Mission Alignment:** By providing a unified platform for AI model development and agentic workflows, AmSC directly supports the Genesis Mission's goal of accelerating scientific discovery through AI integration. It provides the necessary "Information Space Laboratory" described in Section 7, enabling rapid iteration on scientific ideas.
- **Agentic Framework and MCP:** AmSC is designing an **Intelligent Interfaces** layer to support AI agents via **Model Context Protocol (MCP)** servers. This allows an AI agent to "call" a supercomputer simulation tool as easily as a Python function, enabling the fully automated "self-driving" cycles envisioned for future colliders.
- **Collaboration on Use Cases:** To ensure this infrastructure meets the unique needs of particle physics, the HEP community must engage in co-designing specific use cases with AmSC and industry partners. This collaboration will drive the development of features such as low-latency triggers and petabyte-scale data loaders.

8 Conclusion

This whitepaper reflects the collective input, expertise, and ambition of a broad cross-section of the U.S. particle physics community. Contributors spanning many experiments, frontiers, and technical domains have articulated a compelling vision for how AI can transform experimental particle physics: accelerating discovery, enabling new scientific capabilities across the full experimental lifecycle, and empowering a new generation of scientists with the skills needed to lead in an AI-enabled research landscape.

At the same time, this vision is inherently dynamic. AI technologies, experimental programs, and scientific goals will continue to evolve, as will our understanding of where AI delivers the greatest impact. Progress on the projects and ideas outlined here will inform new directions, reveal unanticipated opportunities, and refine existing approaches. Sustaining scientific impact, therefore, requires an iterative strategy that evolves alongside advances in both AI and experimental particle physics. For these reasons, this whitepaper is best viewed as the foundation of a living community vision, rather than a static document. Periodic updates would allow the community to incorporate new insights, assess progress, and re-calibrate priorities. Through a sustained, national-scale collaboration that unites laboratories, universities, and shared AI infrastructure around these Grand Challenges, the community can translate this evolving vision into coordinated action and lasting scientific impact.

References

[1] Hitoshi Murayama, Shoji Asai, Karsten Heeger, Amalia Ballarino, Tulika Bose, Kyle Cranmer, Francis-Yan Cyr-Racine, Sarah Demers, Cameron Geddes, Yuri Gershtein, Beate Heinemann, JoAnne Hewett, Patrick Huber, Kendall Mahn, Rachel Mandelbaum, Jelena Maricic, Petra Merkel, Christopher Monahan, Peter Onyisi, Mark Palmer, Tor Raubenheimer, Mayly Sanchez, Richard Schnee, Sally Seidel, Seon-Hee Seo, Jesse Thaler, Christos Touramanis, Abigail Vieregg, Amanda Weinstein, Lindley Winslow, Tien-Tien Yu, and Robert Zwaska. *Exploring the Quantum Universe: Pathways to Innovation and Discovery in Particle Physics*. June 2023.

[2] LSST Dark Energy Science Collaboration, Eric Aubourg, Camille Avestruz, Matthew R. Becker, Biswajit Biswas, Rahul Biswas, Boris Bolliet, Adam S. Bolton, Clecio R. Bom, Raphaël Bonnet-Guerrini, Alexandre Boucaud, Jean-Eric Campagne, Chihway Chang, Aleksandra Ćiprijanović, Johann Cohen-Tanugi, Michael W. Coughlin, John Franklin Crenshaw, Juan C. Cuevas-Tello, Juan de Vicente, Seth W. Digel, Steven Dillmann, Mariano Javier de León Dominguez Romero, Alex Drlica-Wagner, Sydney Erickson, Alexander T. Gagliano, Christos Georgiou, Aritra Ghosh, Matthew Grayling, Kirill A. Grishin, Alan Heavens, Lindsay R. House, Mustapha Ishak, Wassim Kabalan, Arun Kannawadi, François Lanusse, C. Danielle Leonard, Pierre-François Léget, Michelle Lochner, Yao-Yuan Mao, Peter Melchior, Grant Merz, Martin Millon, Anais Möller, Gautham Narayan, Yuuki Omori, Hiranya Peiris, Laurence Perreault-Levasseur, Andrés A. Plazas Malagón, Nesar Ramachandra, Benjamin Remy, Cécile Roucelle, Jaime Ruiz-Zapatero, Stefan Schuldt, Ignacio Sevilla-Noarbe, Ved G. Shah, Tjitske Starkenburg, Stephen Thorp, Laura Toribio San Cipriano, Tilman Tröster, Roberto Trotta, Padma Venkatraman, Amanda Wasserman, Tim White, Justine Zeghal, Tianqing Zhang, and Yuanyuan Zhang. Opportunities in ai/ml for the rubin lsst dark energy science collaboration. 2026.

[3] US Department of Energy (USDOE). A new era of discovery: The 2023 long range plan for nuclear science. Technical report, US Department of Energy (USDOE), Washington, DC (United States). Office of Science, 10 2023.

[4] National Academies of Sciences, Engineering, and Medicine. *Elementary Particle Physics: The Higgs and Beyond*. National Academies Press, Washington, DC, 2025.

[5] Phiala Shanahan, Kazuhiro Terao, and Daniel Whiteson. Snowmass 2021 computational frontier compf03 topical group report: Machine learning, 2022.

[6] Daniel Anglés-Alcázar, Animashree Anandkumar, Joshua C Agar, Bedrich Benes, Ying Ding, Wei Ding, Lili Du, Jennifer Dy, Baskar Ganapathysubramanian, Krishna Garikipati, Jing Gao, Omar Ghattas, Jane Greenberg, Paul C Hanson, Marti Hearst, Phil Harris, Shirley Ho, Mingyi Hong, Shih-Chieh Hsu, Shuiwang Ji, Anuj Karpatne, Carl Kingsford, Vipin Kumar, L. Ruby Leung, Edgar Lobaton, Madhav V. Marathe, Nirav Merchant, Peetak Mitra, Mark S. Neubauer, Xia Ning, Yuhan Douglas Rao, Balaji Rajagopalan, Xinghua Shi, Jianhui Sun, Brandon Sutherland, Eric Toberer, Wei Wang, Jianwu Wang, and Aidong Zhang. AI to Accelerate Science and Engineering Discovery, 2023.

[7] A Virtuous Cycle: Generative AI and Discovery in the Physical Sciences. `https://mit-genai.pubpub.org/pub/ewp5ckmf/release/2`.

[8] Andrew Ferguson, Marisa LaFleur, Lars Ruthotto, Jesse Thaler, Yuan-Sen Ting, Pratyush Tiwary, Soledad Villar, E. Paulo Alves, Jeremy Avigad, Simon Billinge, Camille Bilodeau, Keith Brown, Emmanuel Candes, Arghya Chattopadhyay, Bingqing Cheng, Jonathan Clausen, Connor Coley, Andrew Connolly, Fred Daum, Sijia Dong, Chrisy Xiyu Du, Cora Dvorkin, Cristiano Fanelli, Eric B. Ford, Luis Manuel Frutos, Nicolás García Trillos, Cecilia Garraffo, Robert Ghrist, Rafael Gomez-Bombarelli, Gianluca Guadagni, Sreelekha Guggilam, Sergei Gukov,

Juan B. Gutiérrez, Salman Habib, Johannes Hachmann, Boris Hanin, Philip Harris, Murray Holland, Elizabeth Holm, Hsin-Yuan Huang, Shih-Chieh Hsu, Nick Jackson, Olexandr Isayev, Heng Ji, Aggelos Katsaggelos, Jeremy Kepner, Yannis Kevrekidis, Michelle Kuchera, J. Nathan Kutz, Branislava Lalic, Ann Lee, Matt LeBlanc, Josiah Lim, Rebecca Lindsey, Yongmin Liu, Peter Y. Lu, Sudhir Malik, Vuk Mandic, Vidya Manian, Emeka P. Mazi, Pankaj Mehta, Peter Melchior, Brice Ménard, Jennifer Ngadiuba, Stella Offner, Elsa Olivetti, Shyue Ping Ong, Christopher Rackauckas, Philippe Rigollet, Chad Risko, Philip Romero, Grant Rotskoff, Brett Savoie, Uros Seljak, David Shih, Gary Shiu, Dima Shlyakhtenko, Eva Silverstein, Taylor Sparks, Thomas Strohmer, Christopher Stubbs, Stephen Thomas, Suriyanarayanan Vaikuntanathan, Rene Vidal, Francisco Villaescusa-Navarro, Gregory Voth, Benjamin Wandelt, Rachel Ward, Melanie Weber, Risa Wechsler, Stephen Whitelam, Olaf Wiest, Mike Williams, Zhuoran Yang, Yaroslava G. Yingling, Bin Yu, Shuwen Yue, Ann Zabludoff, Huimin Zhao, and Tong Zhang. The future of artificial intelligence and the mathematical and physical sciences (ai+mps), 2025.

[9] Website - A Living Review of Machine Learning for Particle Physics. https://iml-wg.github.io/HEPML-LivingReview/.

[10] Matthew Feickert and Benjamin Nachman. A living review of machine learning for particle physics, 2021.

[11] Website - Simulation Based Inference. https://simulation-based-inference.org.

[12] PIONEER Collaboration. European strategy for particle physics update – pioneer: a next generation rare pion decay experiment, 2025.

[13] Website - Seismological Facility for the Advancement of Geoscience (SAGE). https://www.iris.edu/hq/sage.

[14] Website - Geodetic Facility for the Advancement of Geoscience (GAGE). https://www.unavco.org/what-we-do/gage-facility/.

[15] Website - National Ecological Observatory Network (NEON). https://www.neonscience.org.

[16] M. Kagan and L. Heinrich. Presentation - New Frontiers in AI for Fundamental Physics. Based on presentation https://indico.cern.ch/event/1604420/contributions/6760798/.

[17] Website - High Energy Physics - Center for Computational Excellence (HEP-CCE)). https://www.anl.gov/hep-cce.

[18] Website - Institute for Research and Innovation in Software for High Energy Physics (IRIS-HEP). http://iris-hep.org.

[19] National Science Foundation Cooperative Agreement OAC-1836650. https://www.nsf.gov/awardsearch/showAward?AWD_ID=1836650&HistoricalAwards=false.

[20] National Science Foundation Cooperative Agreement PHY-2323298. https://www.nsf.gov/awardsearch/showAward?AWD_ID=2323298&HistoricalAwards=false.

[21] Website - A3D3: Accelerated Artificial Intelligence Algorithms for Data-Driven Discovery. https://a3d3.ai.

[22] National Science Foundation Cooperative Agreement OAC-2117997. https://www.nsf.gov/awardsearch/show-award?AWD_ID=2117997.

[23] Website - NSF AI Institute for Artificial Intelligence and Fundamental Interactions (IAIFI). https://iaifi.org.

[24] Website - A3D3: Accelerated Artificial Intelligence Algorithms for Data-Driven Discovery - Postbac Program. https://a3d3.ai/education-and-outreach/postbac/.

[25] Website - Training to Advance Computational High Energy Physics in the Exascale Era (TAC-HEP). https://tac-hep.org.

[26] Website - Western Advanced Training for Computational High-Energy Physics (WATCHEP). https://watchep.org.

[27] Website - Chicagoland Computational Traineeship in High Energy Particle Physics (C^2 the P^2. https://www.c2thep2.org.

[28] DE-SC-26-016: Request for Information (RFI) on Mobilizing Talent for the Genesis Mission and Developing an American Workforce to Advance Artificial Intelligence (AI) for Science and Engineering. https://sam.gov/workspace/contract/opp/1f6ee2898b724568b4816de787d0b8f6/view.

[29] National Center for Science and Engineering Statistics. Research doctorate recipients' sources of financial support, by broad field of doctorate and sex: 2024. Data Table NSF 25-349, National Science Foundation, 2025.

[30] National Center for Science and Engineering Statistics (NCSES). Doctorate recipients from u.s. universities: 2020. Report NSF 22-300, National Science Foundation, Alexandria, VA, 2021.

[31] Website - DUNE Training ExperienCe Hub (DUNE-TECH). https://dune-tech.rice.edu.

[32] Website - IAIFI Summer School. https://iaifi.org/phd-summer-school.html.

[33] Website - Computational and Data Science Training for High Energy Physics (CoDaS-HEP). https://codas-hep.org.

[34] Website - QuarkNet. https://quarknet.org.