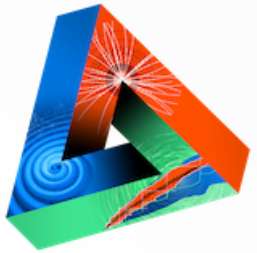
EUCAIF and European Strategy

AI-HEP UK, Liverpool, 17/9/2025

Caterina Doglioni, University of Manchester
Monica D'Onofrio, University of Liverpool

AI and computing in international landscape

- For computing needs (and AI) → JENAA: Joint ECFA-NuPECC-APPEC Activities
  - Started in 2022, **JENA computing** workshops held yearly
  - https://indico.scc.kit.edu/event/3813/
- More specific for AI: EuCAIF → European AI for Fundamental Physics → Conferences in 2024 and 2025 (with some UK representation)
- **Rationale:**
  - European initiative for advancing the use of Artificial Intelligence (AI) in Fundamental Physics.
  - Members are working on particle physics, astroparticle physics, nuclear physics, gravitational wave physics, cosmology, theoretical physics as well as simulation and computational infrastructure.
- 5 Working groups organised around specific themes
- "Fellows" cohort [link] serves as broader organising committee, on top of core coordination group



JENAA

Joint ECFA-NuPECC-APPEC Activities

Aim of the events:
provide a **platform** for establishing new connections between AI activities across various branches of fundamental physics, by bringing together researchers that face similar challenges and/or use similar AI solutions.

EUCAIF: Working group structure and mission

- **WG1: Foundation models.**
  - Mission: Facilitate research on large-scale foundation models for fundamental physics
  - Provide infrastructure, resources, data and models, connect researchers, define problems & metrics
- **WG2: AI-assisted co-design for detectors**
  - AI-assisted co-design of future ground- and space-based detectors - Related to the work of the MODE Collaboration
- **WG3: FAIR-ness and environmental sustainability**
  - FAIR: findable, accessible, interoperable, and reusable principles for data/algorithms, FAIR4ML as part of trustworthy AI
  - See talk this afternoon for environmental sustainability, see also JENA computing workshop talk by K. Lohwasser
- **WG4: ML and AI infrastructure –** Mission: understand status, guide funding and development efforts
  - Scalability and access to HPC Resources, Data Infrastructure and Infrastructure for Distribution, Integration of AI/ML into Existing Experimental Workflows, Large and Small Scale Models, Benchmarking and Standards
- **WG5: Building-bridges - community, connection**
  - Establish EuCAIF as a new brand for AI in physics (independent of CERN), eventually bringing in new grants
  - Aggregate and provide an overview of already existing structures and networks, put this on professional website

Overall, some engagement from UK HEP community (C. Doglioni in WG3), 8 UK fellows so far, but it can still grow

EUCAIF activities

- **WG meetings on indico [link]**
- **Organisation of yearly conference**
  - Two editions so far (Amsterdam, Cagliari)
  - Very good attendance, lots of space for Early Career Researchers to present / discuss
- **Gathering people with similar interests towards collaborative grant applications**
  - Not completely straightforward (only one “public” call)
  - A few applications in the works since the last EuCAIFCon
- **Whitepaper as input to European Strategy (see next slides)**

Inputs for European Strategy PP update

- 2 main inputs submitted and considered for the Briefing Book
  - AI-RDs: A EuCAIF proposal to structure AI research in Particle Physics
  - EuCAIF Recommendations for Scaling AI Capabilities in Particle Physics (mirror closely UK ones)

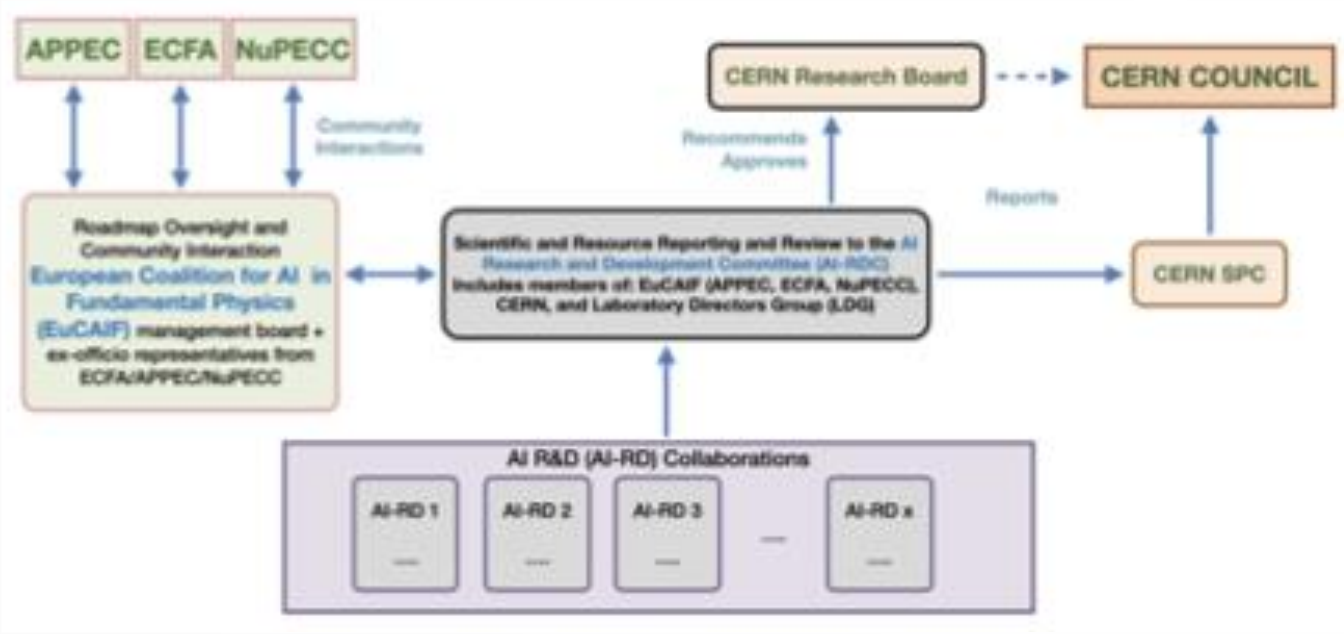


Figure 1. Structure of the proposed AI Research and Development collaborations (AI-RDs), the AI Research and Development Committee (AI-RDC), and integration with the European Coalition for AI in Fundamental Physics (EuCAIF), CERN, the European Committee for Future Accelerators (ECFA), the Nuclear Physics European Collaboration Committee (NuPECC), and the Astroparticle Physics European Consortium (APPEC).

- **AI for Detector and Accelerator Control:** AI optimizes beam tuning, accelerator performance, experimental design, calibration, predictive maintenance, and system monitoring.
- **AI for Detector Optimization:** AI-based workflows to optimize the design of new detectors, exploiting techniques like differentiable programming and reinforcement learning to maximize detector performance while designing AI-friendly detector architectures that would facilitate the use of AI techniques for simulation, reconstruction, and other computing-heavy central tasks.
- **AI for Event Reconstruction:** AI-driven reconstruction algorithms, from tracking and calorimetry from detector hits to particle flow, can give sizable improvements to downstream data analysis steps, while improving the real-time processing in trigger systems. Further development in this direction could lead to the development of global-event end-to-end approaches with foundation models.
- **AI for Analysis:** AI-driven applications to data analysis, like signal-background separation, optimal variable definition, and likelihood-free inference approaches transform data analysis. They lead to an optimal understanding of the data, comprehensive uncertainty quantification and a statistical interpretation. An promising direction is representation learning with symmetry-aware architectures and foundation models.
- **AI for Generation and Simulation:** AI accelerates simulations with generative models, surrogate modeling, and AI-enhanced Monte Carlo methods. It enables, for instance, fast and precise evaluation of precise theory predictions, event generation, and detector simulation. Simulation-based inference and generative networks relate this topic to AI for Analysis.
- **AI for Theory:** AI assists in high-dimensional parameter optimization and numerically challenging lattice computations. It helps solve differential equations and allows symbolic regression, related to fundamental physical laws. These applications are focused on theory, but for instance parameter optimization, sampling, and Monte Carlo integration reach much further.
- **AI for HPC Usage:** AI tools will facilitate the exploitation of highly parallelized, large-scale GPU computing. Using the full parallel computing power of GPUs is a challenge for classic simulation methods, for instance Monte Carlo methods. In contrast, AI-methods like surrogates and neural sampling are perfectly suited for highly parallelized architectures.
- **AI for Documentation and Education:** Conversational AI to transform experiment-specific and broader physics communication and training. Goals include concept understanding, accelerating learning, and idea development. To fully realize this potential, specialized large physics-language models need to be aware of domain-specific

EUCAIF recommendations

- Very close to ours..

- (R1) Convene dedicated discussions with national research groups and funding bodies to assess and compare the feasibility of a **centralized large-scale GPU facility versus federated and hybrid high-performance computing (HPC) infrastructures**, supported by working groups developing detailed implementation plans for both options, with the aim of accelerating the deployment of a scalable AI infrastructure.

- (R2) Establish a **scalable data infrastructure** initiative by creating shared repositories and tools, and **developing platforms for distributed workloads**. These efforts need targeted funding programs and a concrete community-driven structure to ensure widespread adoption and collaboration in AI research.

- (R3) Encourage funding to **transition AI-driven R&D activities into production-ready applications** within established experimental workflows, focusing on adopting best practices to achieve practical, scalable improvements without requiring a complete system overhaul.

- (R4) Allocate **dedicated funding to establish and support specialized Machine Learning Operations (MLOps) personnel** to streamline the integration and ensure the sustainable maintenance of AI models within production workflows. This effort should encompass the development of community-wide standards, tools, and platforms to effectively manage the entire lifecycle of machine learning models.

- (R5) Invest in the **creation of "science Large Language models (LLMs)"** tailored to the unique challenges of fundamental physics and science, balancing the use of commercial tools for general tasks with specialized models for domain-specific needs. This requires dedicated funding, access to large-scale GPU infrastructure, and collaborative frameworks to enable transparent, efficient, and impactful AI solutions.

- (R6) Establish dedicated funding schemes and a collaborative structure to develop community-driven **foundation models trained on domain-specific data to learn meaningful representations serving a large variety of downstream tasks**. This effort should identify representative benchmarks, extendible in complexity and realism by integrating both synthetic and real-world data to address domain-shift issues, leverage physics-informed augmentations, ensure models are rooted in scientifically relevant tasks, and foster automation, explainability and interpretability to accelerate AI advancements in the field, and to develop a well-defined AI demonstrator for the wider AI community.

EUCAIF recommendations (2)

- (R7) Establish a dedicated effort to **develop and maintain extensible benchmarks for various AI tasks in fundamental physics**, such as event classification, parameter inference, tracking and anomaly detection. Support efforts to encourage researchers to share well-documented surrogate models to promote reusability and collaboration to drive innovation and standardisation in this area.
- (R8) Investigate and adopt benchmarks that are suitable for fundamental sciences to raise **awareness of the environmental impact** of large AI models. Consider collective mitigation strategies such as optimising widely used frameworks and models and their interfaces to existing software frameworks, as well as individual strategies that lead to minimal/acceptable performance loss. Cooperate with infrastructure and computing sites to minimise carbon costs of compute-intensive AI tasks.
- (R9) Develop activities aiming to **integrate FAIR compliance into publication criteria and practices**, recognise and incentivise the FAIR compliant work in policy and funding measures as well as career progression, build community awareness through training and collaboration, and support the development of technical tools and standards to facilitate the adoption of the FAIR principles.
- (R10) Fund the **development and organization of practical training courses and summer schools** to equip researchers with the skills to implement open research and reproducibility requirements, incorporating examples and industry perspectives. Facilitate partnerships with industry to sponsor training events and provide placements for early-career and senior researchers, enhancing their AI and data science expertise while fostering connections between fundamental science and commercial applications.
- (R11) **Establish interdisciplinary research initiatives** that bring together physicists, AI specialists, software engineers, HPC experts, and potentially experts from other related fields, to tackle large-scale projects. Provide dedicated funding to support **cross-domain knowledge transfer** through workshops, training programmes and open source collaboration. Invest in shared repositories and computing platforms to enable data sharing, modelling development and collaboration between different disciplines.
- (R12) **Establish and support a dedicated organisational structure** to coordinate strategic investments in AI for fundamental physics to accelerate the development and deployment of innovative AI technologies tailored to the specific challenges of the field. Existing initiatives like the European Coalition of AI for Fundamental Physics (EuCAIF) can serve as a model for such efforts.

The next decades…

- **The Next 10 Years (2025–2035):** AI is already mature but much more to be done:
  - Classification and Tagging
  - Offline Reconstruction algorithms
  - Simulation and Generative Modelling
  - Triggers and Real-Time Systems
  - Inference and Unfolding

In BB for collider physics but true across the board..

- **Beyond 2035** [according to the ES review] →AI for next-generation colliders/experiments:
  - AI will be at the very heart of experimental design and operation → traditional workflows of collection, storage, and offline analysis no longer be viable → AI could be embedded directly into detectors and data pipelines
    - Development of intelligent frontends (extractors integrated on ASICs, with early prototypes already demonstrated in CMOS technology)
    - Foundation models
    - Large language models → Large Physics models (LPMs)
    - Differentiable programming → geometry and readout of detectors optimized through gradient descent, directly targeting physics performance – merging detector development and physics optimisation
  - **Is this perhaps even sooner that 2035?**
- **Add on: The quantum frontier:** quantum-AI or AI for quantum

Some (draft) conclusions (from ES – very close to ours)

- Invest in both **infrastructure** and **people**
- Establish a sound and **solid software ecosystem**, to ensure comparability and reproducibility, i.e. share benchmarking datasets and evaluation metrics, as well as fast pacing
- Nurture the “human capital” → dedicated training programs, establish attractive career paths: prepare the next generation of particle physicists “fluent” in AI and ensure they have a future in the field
- Exploit the ‘**sustainability route**’ – responsible energy consumption for AI algorithms is a must (but also: target efficient and sustainable software for our future experiments pass through AI)

Back up

Some slides from past meeting summary …