

DATA SCIENCE AND AI

Monica, Carsten
8/11/2024





Rationale

- Data science underpins most of the research the department is involved in.
- AI and machine learning approaches have become increasingly vital to a very wide range of our physics programme
 - Can enhance almost all aspects of our experiments and some of theory: Object reconstruction, event selection, simulation, event generation, data quality, detector control/monitoring, coding, trigger, hardware design, documentation....
 - Many examples where applying AI has been truly transformational
- Activities are on-going in all clusters with similarities in approach and objectives (online and offline data analysis for PP, application to detector developments and data analysis for NP, application to medical science in CMP, applications to accelerator physics in AS, analysis of usage of AI in HE in PER)
- The LIV.INNO doctoral centre (and LIV.DAT before) is an excellent basis for potential cross-disciplinary research activities on AI, supporting research in astronomy, accelerator, nuclear, theoretical and particle physics.
 - the first real large-scale cross-cluster collaboration in our department (and other departments), and focus on using similar skills sets to enable a broad and multi-area research program.

Moving forward, it is desirable to **capitalise on the diverse expertise within the department**, increasing research income exploiting cross-collaboration, investing in collaborations with industry partners and/or national facilities, continuing/establishing dedicated training opportunities for students.



Potential targets

- Future specific targets for us include (but are not limited to):
 - **Improve networking:** national and international labs, collaborations with other universities (Hubs) , continue trainings and exploit CDT opportunities further
 - **Identification of opportunities for large bids** and establishment of multidisciplinary cases that can cover a variety of applications for which common methodologies can be deployed;
 - Access to **computing resources** adequate for future challenges: exascale computing
 - **take part to consortia**, i.e. with computer science and with industry partners in the field;
 - **establish a PGT programme on data science and AI**, targeting students in Physics and Mathematical Science – a proposal has been made with part of the provision being offered by Physics and part by Math (lead proponent), fill the existing gap in Higher Education in the North area on data science.
 - **AI in education:** some studies on-going, could be further explored and expanded.



National landscape: exascale computing and AI strategies

- UKRI recognized important of AI since few years
 - <https://www.ukri.org/who-we-are/our-vision-and-strategy/strategies-and-reviews/ai-review-transforming-our-world-with-ai/>
- STFC and EPSRC – strong focus on data science and AI
 - EPSRC-driven existing hubs: <https://www.ukri.org/news/100m-boost-in-ai-research-will-propel-transformative-innovations/>
 - STFC AI Strategy in preparation: *position and ambitions to advance AI across the breadth of scientific communities, large-scale national science facilities, and industrial partners. It also defines STFC's role within the national AI and data R&I ecosystem.*
 - [EXATEPP: https://www.hartree.stfc.ac.uk/work-with-us/projects/exatepp/](https://www.hartree.stfc.ac.uk/work-with-us/projects/exatepp/) Theoretical and Experimental Particle Physics at the Exascale Frontier (ExaTEPP) at Hartree Centre
- AI in PP plans – applicable across all research fields

In order to realise the full benefits of AI, we will need to act to support the growth of the UK's AI research and innovation capabilities, building on strong foundations. This is likely to require substantial investment over a sustained period.



From AI for Particle Physics

- Plans and challenge common to all

Progressing Further?

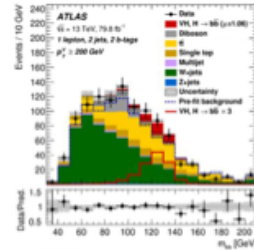
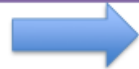
- AI has potential for further transformative change
 - Pushing the boundary:
 - Many (all?) students use AI tools to write code
 - Could envisage AI models that assist in the full analysis chain
 - *Focus more on physics rather than code frameworks?*

```

def calculateMET(event):
    # Calculate MET from event objects
    # ... (code omitted) ...
    return MET
    
```

ChatGPT: Write a function to calculate MET

ChatHEP: 'Produce a data/MC comparison plot with latest theory/exp corrections for H->bb'



- Feed raw data into Large Language Models which can reconstruct objects
 - *Solve CPU issues?*
- Auto-encoders that can compress our data with little loss of information
 - *Assist with data storage issues?*
- Marking all our exams/course work.....

Enabling AI means

- Focus on challenges, barriers and opportunities in:
 - **Software, hardware and ML-Ops**
 - **Skills/Training and capacity building**
 - **Knowledge exchange and wider engagement**
- This is the first exploratory workshop
 - Aim is to collect input on above areas and any additional topics missed
 - Decide on next steps for this initiative
 - Make links between AI focussed members of community

<https://indico.cern.ch/event/1450122/>



International landscape: JENAA initiatives on AI

CHALLENGING
INSPIRING
AMBITIOUS
SPIRITED
INSPIRING

- ▶ Detector R&D always in the remit of ECFA – DRDs formed and organised
- ▶ For computing needs (and AI) → JENAA: Joint ECFA-NuPECC-APPEC Activities
 - ▶ Started in 2022, a first **JENA computing** workshop was held on June 12-14, 2023 (Bologna)
 - ▶ <https://indico.scc.kit.edu/event/3813/>
- ▶ More specific for AI: EuCAIF → European AI for Fundamental Physics → Conference in 2024
 - ▶ EuCAIFCon2024 <https://indico.nikhef.nl/event/4875/>
- ▶ **Rationale:**
 - ▶ new 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.

e-group: eucaif-info@cern.ch

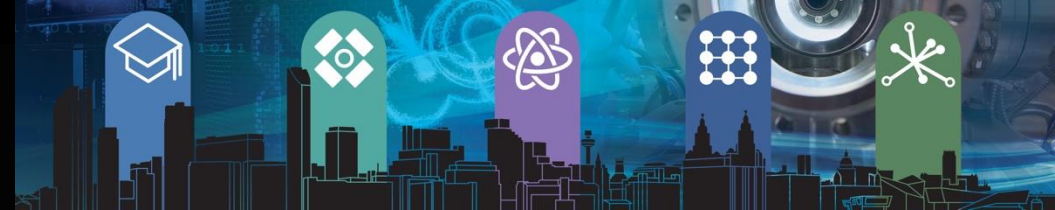


Aim of the event:

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.

5 working groups:

- ▶ WG1: Foundation models
- ▶ WG2: AI-assisted co-design for detectors
- ▶ WG3: FAIR-ness and sustainability
- ▶ WG4: ML and AI infrastructure
- ▶ WG5: Building-bridges - community, connections and funding

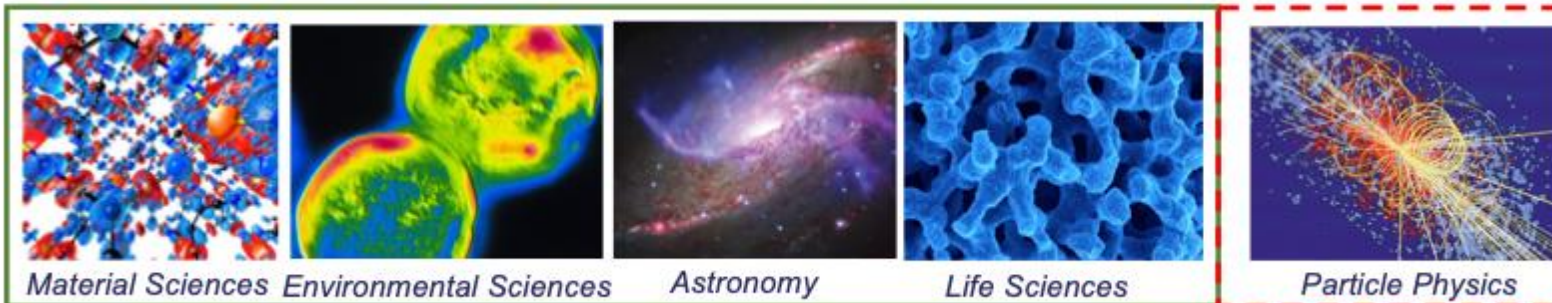


Unlock potential of national facilities

- Example: SciML ([talk](#))

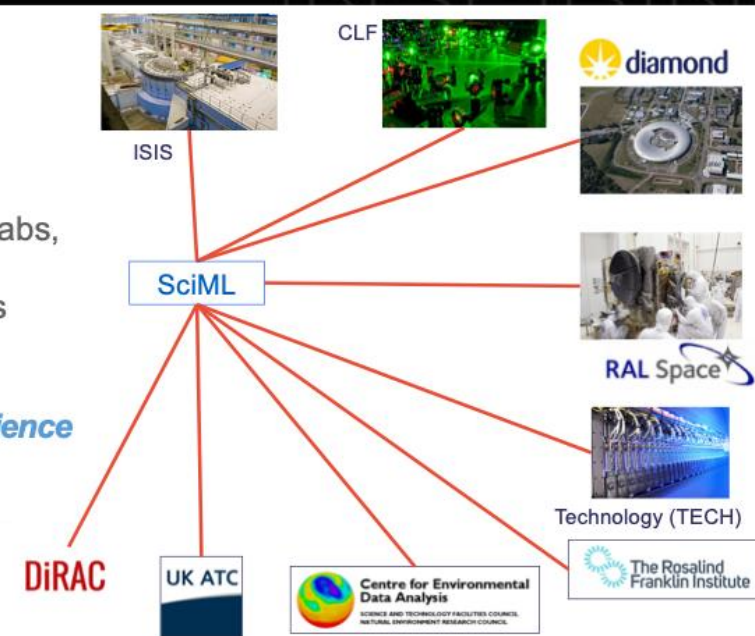
Jeyan Thiyagalingam
Rutherford Appleton Laboratory

Domain Coverage



Our Role

- Work with national labs, facilities and STFC funded programmes
- Underpin the AI capabilities
- Our Remit: *AI for Science*





Training of personnel: MSc and PhD programmes

- Overarching importance of training skilled personnel for development of cutting-edge technologies
- Excellent programme in place for UG and PhD, underpinning all our research / impact activities

LIV.INNO: Major centre for doctoral training (up to 80 PhD students) in AI/data intensive science, encompassing accelerator, particle and nuclear physics.

- Follows on the successful LIV.DAT CDT program also supported by STFC (2019-2022)
- Through LIV.INNO and beyond, more than 50 partnerships with industry at different levels
- Within the program, dedicated events and schools
E.g. July 16-19: STFC School on Data intensive Science:
<https://indico.ph.liv.ac.uk/event/1639/>





Aim of this meeting

- Identify areas where expertise can be shared to achieve specific goals
 - Research grants and collaborations
 - Facilitate access to resources
 - Enhance training pipelines (PGT, CDT)
 - More ..?

CHALLENGING
INSPIRING
AMBITIOUS
SPIRITED
INSPIRING
CHALLENGING
AMBITIOUS
SPIRITED



Discussion topics

- **Improve networking:** national and international labs, collaborations with other universities (Hubs) continue trainings and exploit CDT opportunities further
 - But also within our own institution
 - <https://www.liverpool.ac.uk/research/digital/>
 - <https://www.liverpool.ac.uk/digital-innovation/>
 - Collaborations with computer scientists at Liverpool
- **Access potential income, apply to research grants:**
 - **Identification of opportunities for large bids** and establishment of multidisciplinary cases that can cover a variety of applications for which common methodologies can be deployed;
 - **take part to consortia**, i.e. with computer science and with industry partners in the field;

CHALLENGING
INSPIRING
AMBITIOUS
SPIRITED
CHALLENGING
INSPIRING
AMBITIOUS
SPIRITED



Discussion topics

- Access to **computing resources** adequate for future challenges:
 - Share experience in applying for national computing resources
- But also, **share tools and knowledge**
 - University/school infrastructure for hosting code e.g. github/lab, jupyter server (Carl)
 - Along the same lines, repository for tools (notebooks?) that can boost capability to apply acquired knowledge across the board (MDO)

CHALLENGING
INSPIRING
AMBITIOUS
SPIRITED
INSPIRING
CHALLENGING
AMBITIOUS
SPIRITED

Discussion topics

- **Training: future of CDTs**
- **establish a PGT programme on data science and AI**, targeting students in Physics and Mathematical Science – a proposal has been made with part of the provision being offered by Physics and part by Math (lead proponent), fill the existing gap in Higher Education in the North area on data science.
 - Consider also MPHYS “Physics with AI” ?
- **AI in education**



CHALLENGING
INSPIRING
AMBITIOUS
SPIRITED
INSPIRING
AMBITIOUS
SPIRITED