International PostExascale

Workshop Series

InPEx 2025 workshop - 14-17 April 2025
Pete Beckman, Mohamed Wahib, Thomas Moreau

AI and HPC
Recap from Sitges

Accelerate AI in Science sessions

- 7 oral presentations covering a very wide range of applications and issues/aspects of AI4Science
    - Build/specialize Foundation models for Science, data, open/reproducible models, evaluation
    - Domain-specific models, driven/tailored by physics/application
    - HPC4AI: how we can improve the SW stack to get faster/better/cheaper LLM/FM training
- ...

What is different today in Post-Exa world? Challenges :

- Industry
  - We cannot compete with industry
  - Where we (HPC) can overlap with industry and where to take advantage is not clear?
    - It is difficult to imagine **scientific foundation models** that *do not overlap* with industry
  - Are there new kinds of collaboration with industry available?
  - Do we have to focus on scientific niche, where our community can have a clear added value ?

- AI-in-science governance
  - What role do we expect to play in AI governance and regulations (w.r.t. AI-in-science, not consumer-facing AI)
  - A risk of heavy regulations that could limit
  - Improving AI evaluation for science can inform risk/danger/harm → input to regulations

Challenges :

- Ethical use of knowledge and Science
  - Using open science and open data, as an opportunity to set up guidelines, principles
  - How to evaluate/validate the information generated by AI, and AI solutions in general
  - How to evaluate/validate the robustness of AI models

- How do we engage with the application community
  - Sharing data/helping them to prepare/curate data so as to have AI-ready is key but not everything …
  - Gaining the trust of the application communities is central
    - Application-centric validation ?
    - Physics-informed AI solutions ?
    - Workflows to ease testing new models in domain-specific contexts

Challenges :

- What are the best architectures for Post-Exa hybrid AI / traditional HPC ?
    - Coupling of traditional HPC with AI to develop new models for science
    - Modeling and simulation are critical to feed the development of new models

- Data
    - Openness is paramount in science. How to assure safety.
    - AI-ready data sharing at the international ?

Main Actions (expected):

- Continue improving understanding of how AI4Science transforms the post-Exascale blueprint
- International visiting researcher and student programs exist: we can strategically leverage these paths to advance InPEx
- Shared *scientific* performance evaluation of AI models
- Concrete actions on adapting existing workflows/tools (e.g. HuggingFace) to quickly test/benchmark models for specific data without the pain to re-building them.
    - Dedicated international working group
    - Pilot users to for testing/evaluating
- Collaborate and build scientific “AI Ready” data sharing for AI training and validation

2 Breakout sessions for InPEx 2025 around AI:

- **Hybrid workflow in AI and HPC**
  - Organizers: M. Wahib, J.-P. Villote, T. Moreau
  - Focus on Benchmarking hybrid AI/HPC workflows
  - **Goal:** Identify use-cases for first benchmarks and concrete steps to produce them.
  - **Format:** working group

- **GenAI for sciences**
  - Organizers: A. Buttari, T. Moreau, P. Beckman
  - Focus on the HPC tools necessary to unlock AI for sciences
  - **Goal:** identify the impact of paradigm shift on high performance scientific computing on architecture, code, data, and workflows
  - **Format:** unconferencing style