International PostExascale

Workshop Series

InPEx 2025 workshop – April 15-17, Japan

IAI and HPC : sharing AI-centric benchmarks of hybrid workflows

J.P. Vilotte (CNRS),M. Wahib (R-CCS), T. Moreau (INRIA), F. Capello (ANL)

InPEx

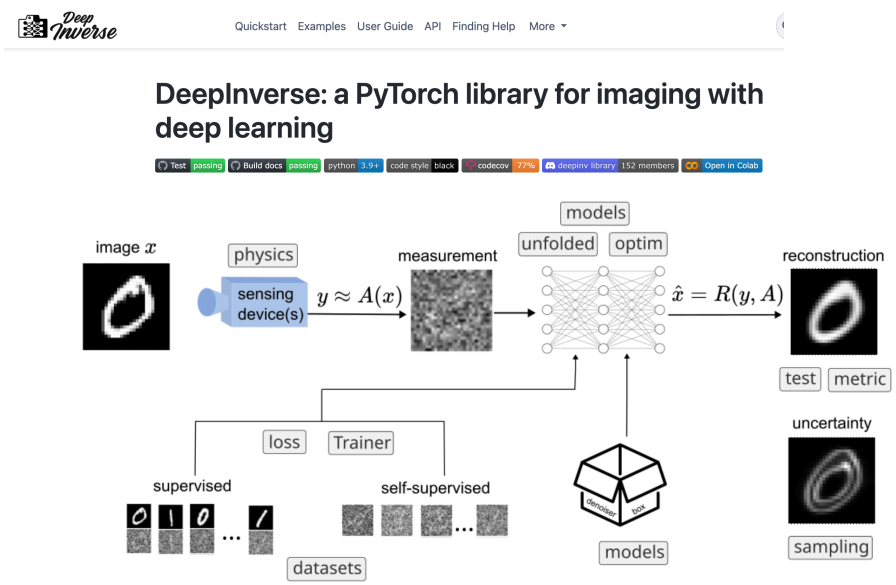
International Post-Exascale Initiative

Context:

- Enhanced importance, use and performance of AI-coupled HPC workflows for a variety of science scenarios
- To unlock fast progress on HPC/AI coupling : need to measure progress
- Opportunity in InPEx: Benchmarks only work if everyone agrees on the evaluation

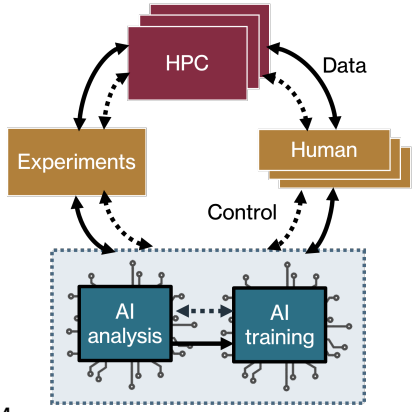
Goals of the session:

- Frame the definition and design of benchmarks for end-to-end AI-coupled HPC/HPDA workflows
- Which common problems/grounds can be identified
- What workflows would benefit from such benchmarks
- Identify applications at various scales that could be included
- Reflect on the evaluation methodologies and possible metrics



Execution Motifs of AI-coupling HPC/HPDA workflows

	Interaction	Coupling	Scope
	Data flow Control flow Human interaction	Concurrency Dynamism Federation	
AI/HPC Workflows	MOTIF		
	Implementation		
	Performance characteristics		



Brewer et al, 2024

First break out session articulated around

1. What are interesting AI-coupled HPC/HPDA workflows and applications we should target?
2. What are the programming and execution models used for this workflow?
3. Are you using workflow management frameworks and which one ?
4. What are the challenges and bottleneck in the execution of this workflow?
5. How an end-to-end evaluation of the workflow should be designed?
6. What are the data available (simulation, experiment, syntheric) for the benchmark?
7. What are the interesting metrics to be be considered?
8. What can we share (data/methodology) and collaboration tools ?

Form discussion groups of ~8

Discuss the questions and come up with use case candidates for the benchmark

The goal is to have concrete examples for the answers.

Inputs of each group in the shared document
https://tinyurl.com/inpex-bench

Presentation of the some results of the discussion

Most of the current benchmarking efforts consider an application or model in isolation: there is a gap in incorporating workflows that help measure end-to-end workflows' performance and the crucial components to accelerate the workflow.

Understanding the audience/stakeholder and motivations to engage in a benchmark exercise

Identified difficulties:

- AI-coupled HPC/HPDA workflows often used by one team and not often beyond the initial use cas
- Portability challenge for complex workflow
- Big question about the scales of the benchmark, with the notion of scaling laws.
- Not one metric to rule them all, but we need diverse metrics

Investigate benchmarks (input-output) with mini-apps representative of execution motif and the actual production runs (to further discussed this afternoon)

- identify common pain points across different workflows
- Importance of scalability, input/output specifications, and the ability to test with various software versions
- Must be representative of actual applications while allowing for flexibility in implementation.