Relevance of LSTM & GRU Models in CoE RAISE

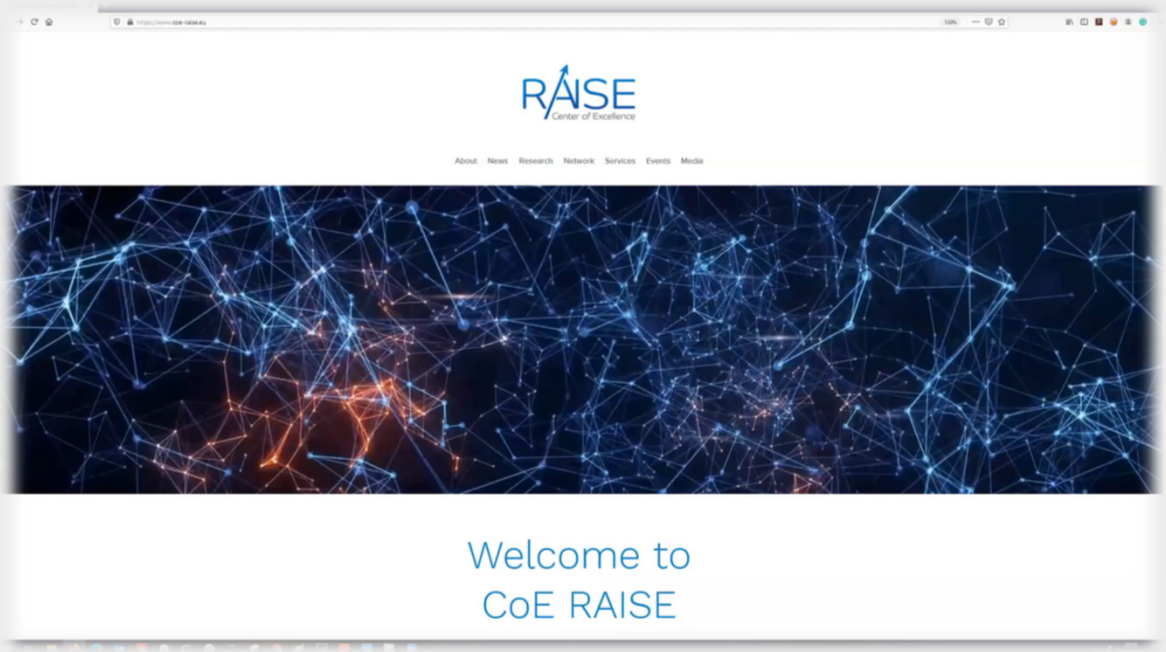
Prof. Dr. – Ing. Morris Riedel
School of Engineering & Natural Sciences, University of Iceland, Iceland
Juelich Supercomputing Centre, Forschungszentrum Juelich, Germany

2022-10-31, RAISE CoE Training LSTM & GRU Models, Online

@ProfDrMorrisRiedel @Morris Riedel @MorrisRiedel @MorrisRiedel
https://www.youtube.com/channel/UCWC4VKHmL4NZgFfKoHtANKg morris@hi.is

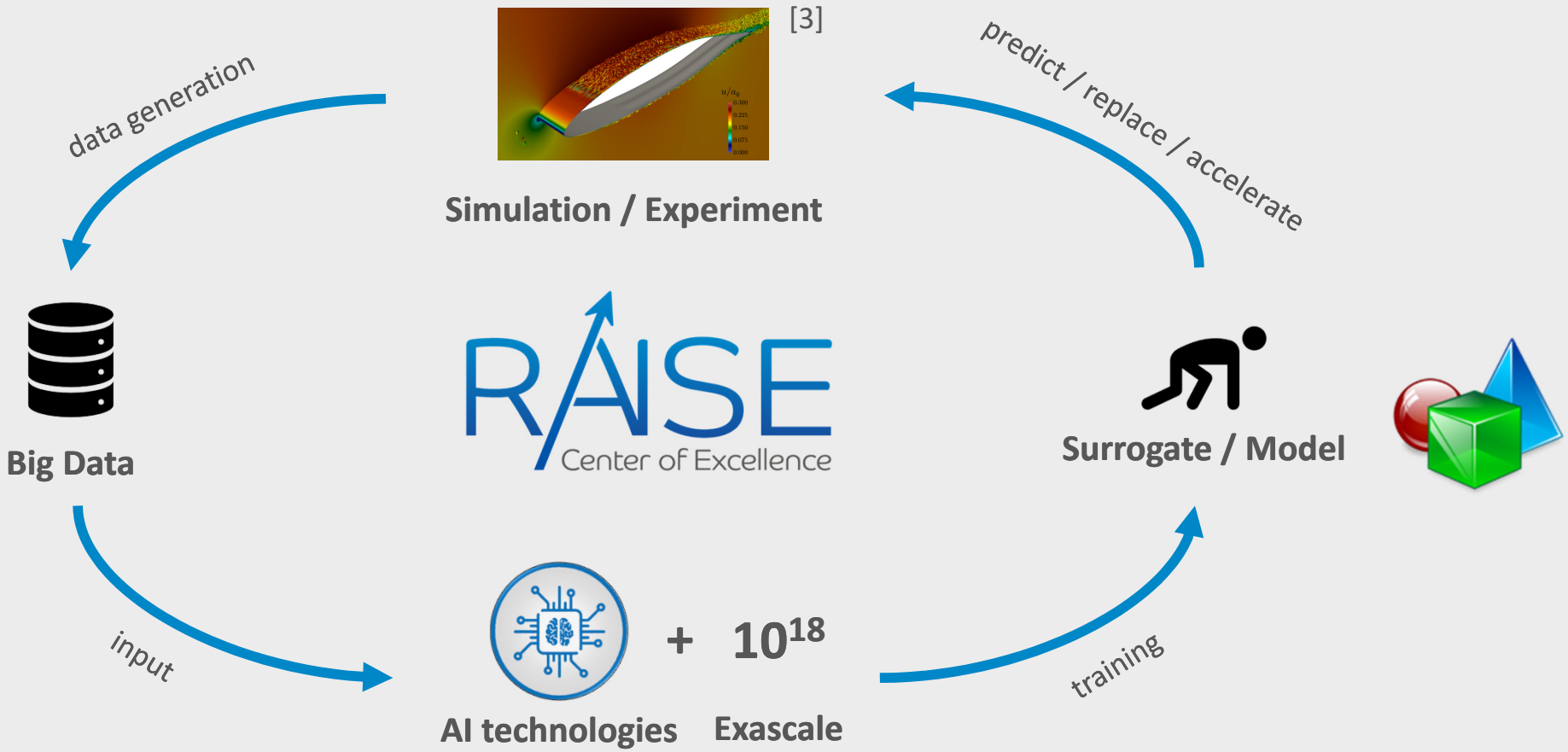
IHPC National Competence Center (NCC) for HPC & AI in Iceland

CoE RAISE Web Page & More Information



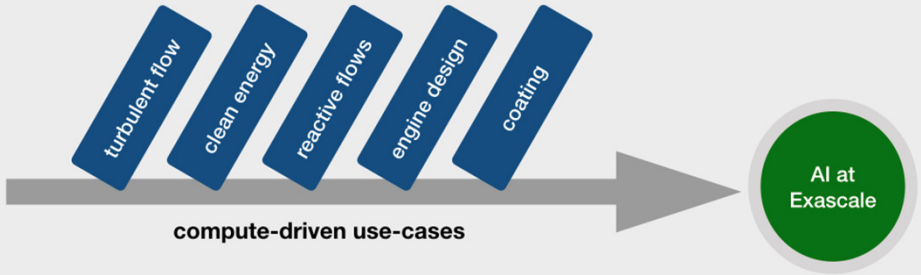
https://www.coe-raise.eu

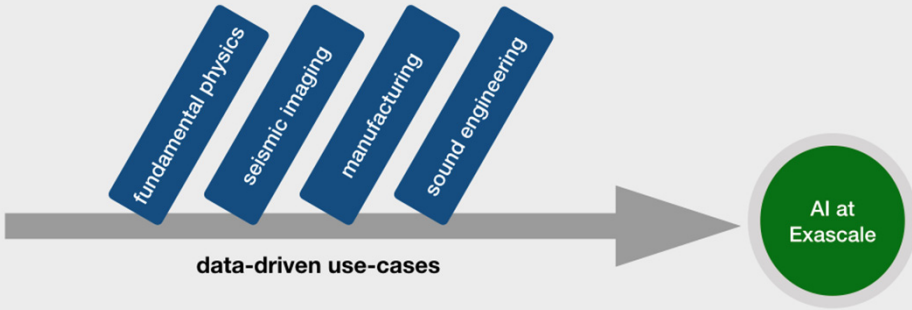
CoE RAISE – Motivation & Approach

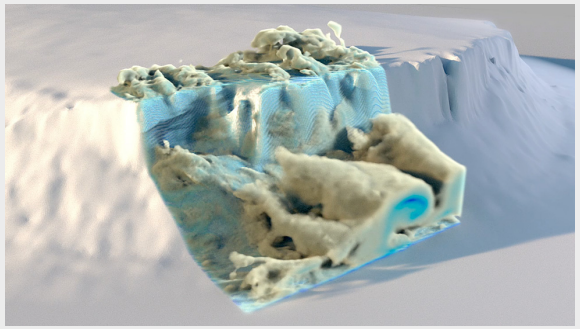


Use Cases in CoE RAISE

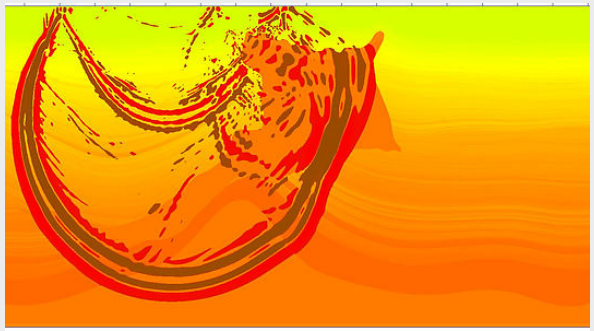
- Two kinds of use cases:





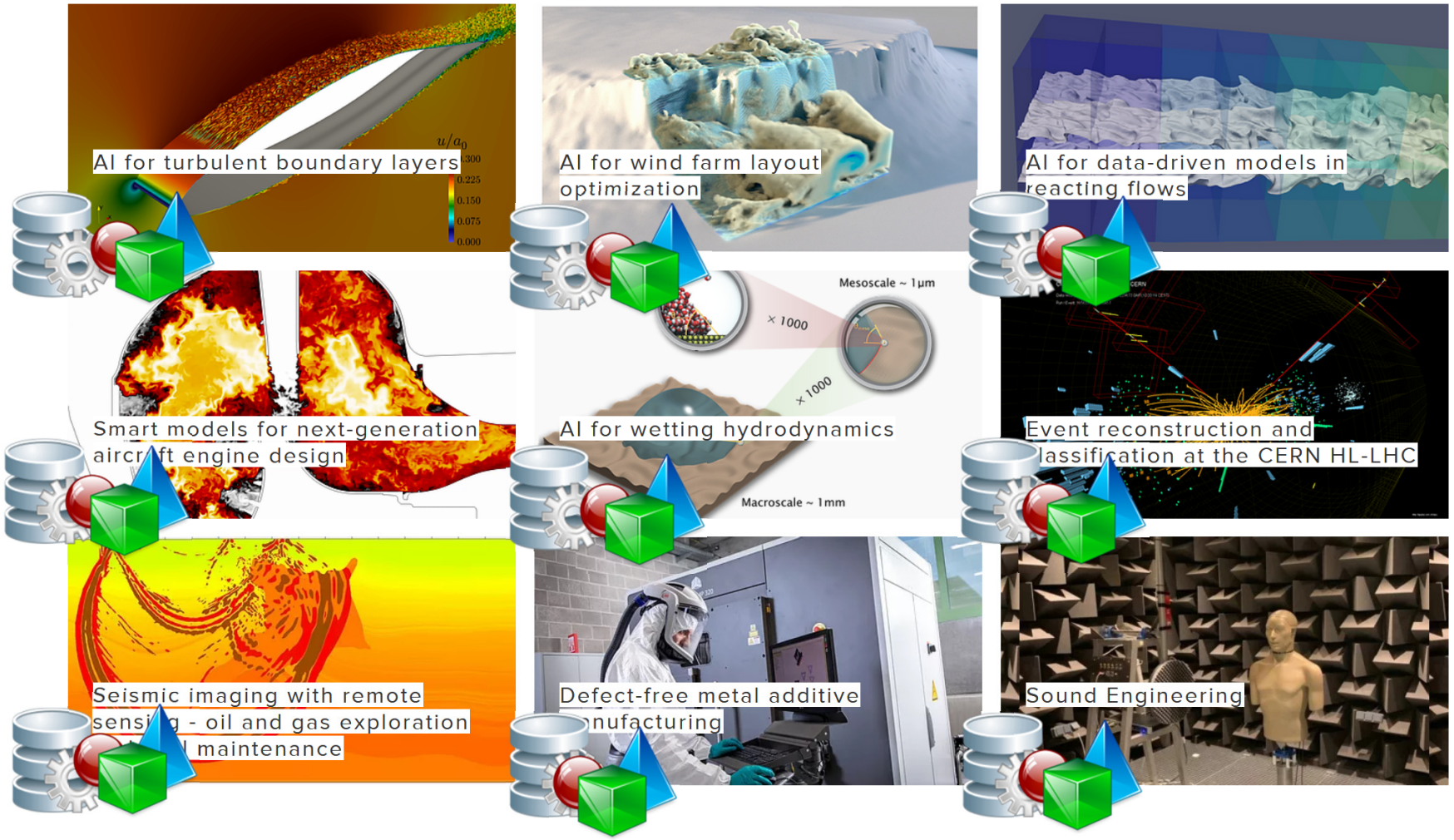


Example from use case ”AI for wind farm layout”: Turbulence generated by a cliff on Bolund Island, Denmark.



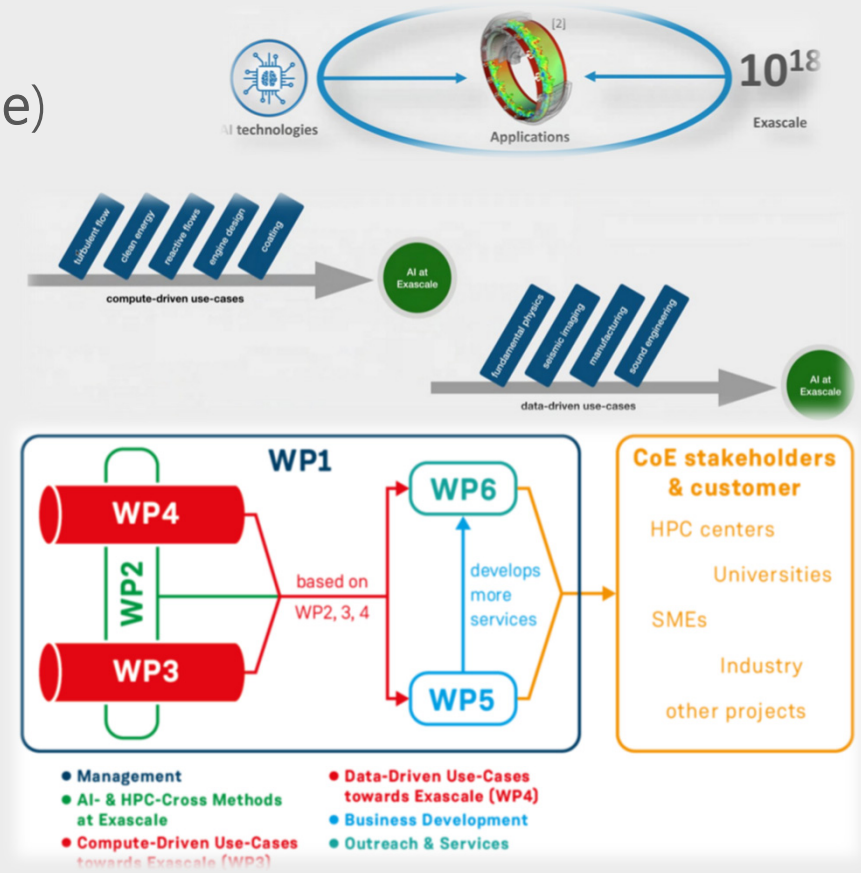
Example from use case ” Seismic imaging with remote sensing - oil and gas exploration and well maintenance”: Snapshot from a wavefield.

Compute- and Data-driven Use Cases – Data & Modeling

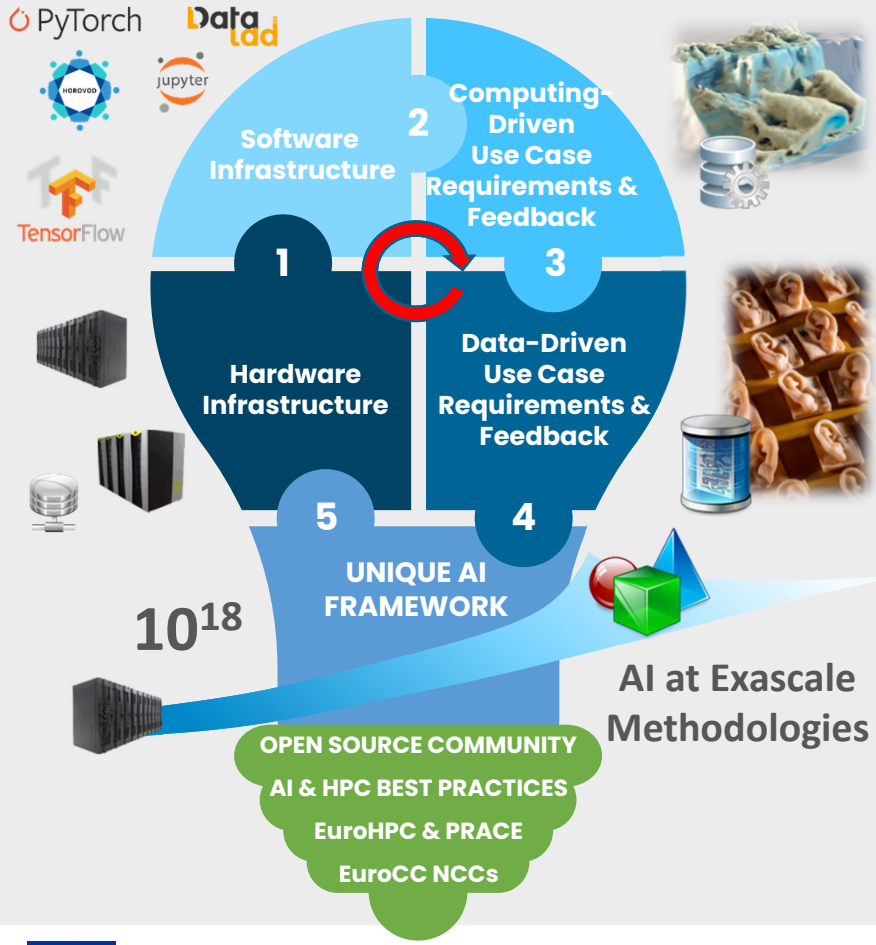


WP2 – AI- & HPC-Cross Methods at Exascale in a nutshell

- WP3 (Compute-Driven Use-Cases towards Exascale)
- WP4 (Data-Driven Use-Cases towards Exascale
- Developments in these WPs will be supported by the cross-linking activities of WP2
  - E.g. scaling machine & deep learning codes with frameworks like Horovod/Deepspeed
  - E.g. introduction to new AI methods such as Long-Short Term Memory (Time series)
  - E.g. data augmentation approaches
  - E.g. benchmarking HPC machines and offer also pre-trained AI algorithms (i.e., transfer learning)
  - E.g. offer neural architecture search methods for hyperparameter – tuning in semi-automatic way



Towards AI & HPC at Exascale with CoE RAISE Results



Hardware Infrastructure

Prepare & Document available production systems at partners' HPC centers

Examples: JUWELS (JUELICH), LUMI (UoICELAND), DEEP Modular Prototypes, JUNIQ (JUELICH), etc.

Software Infrastructure

Prepare & Document available open source tools & libraries for HPC & AI useful for implementing use cases

Examples: DeepSpeed and/or Horovod for interconnecting N GPUs for a scalable deep learning jobs

Computing-driven Use Cases Requirements & Feedback

Use cases with emphasize on computing bring in co-design information about AI framework & hardware

Examples: Use feedback that TensorFlow does not work nicely, so WP2 works with use cases on pyTorch

Data-driven Use Cases Requirements & Feedback

Use cases with emphasize on data bring in co-design information about AI framework & hardware

Examples: Deployment blueprint by using AI training on cluster module & inference/testing on booster

→ UNIQUE AI FRAMEWORK

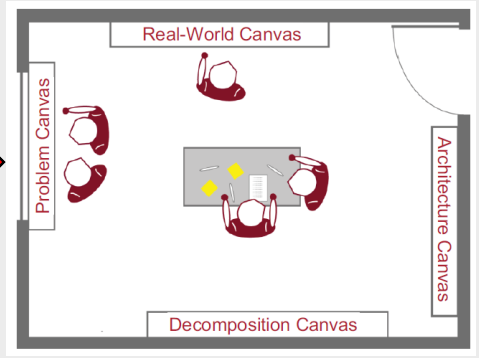
Living design document & software framework blueprint for HPC & AI also with pretrained AI models

Selected Techniques to Identify Cross-Methods for HPC & AI

- Fact Sheets
  - Foster initial understanding
  - Living document & each Fact Sheet per WP3/WP4 Use Case
  - *(Experience from many other EU projects)*
- Selected Contents
  - Short Application Introduction
  - Clarify Primary Contacts
  - Codes/Libraries/Executables
  - HPC System Usage Details
  - Specific Platforms & 'where is what data'?
  - Machine/Deep Learning Approaches of Interest

initial steps

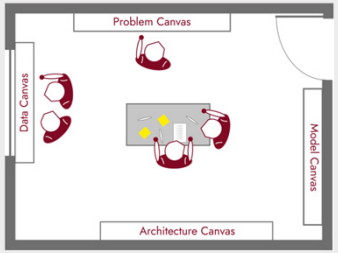
Driven by
Prof. Matthias Book &
Prof. Helmut Neukirchen

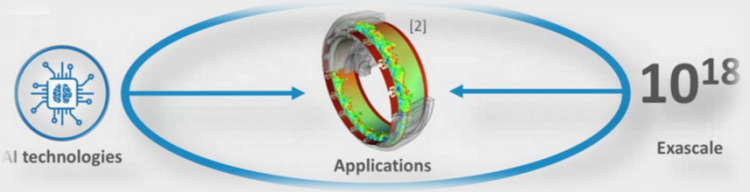


interaction room process

HPC Systems Engineering in the Interaction Room Seminar

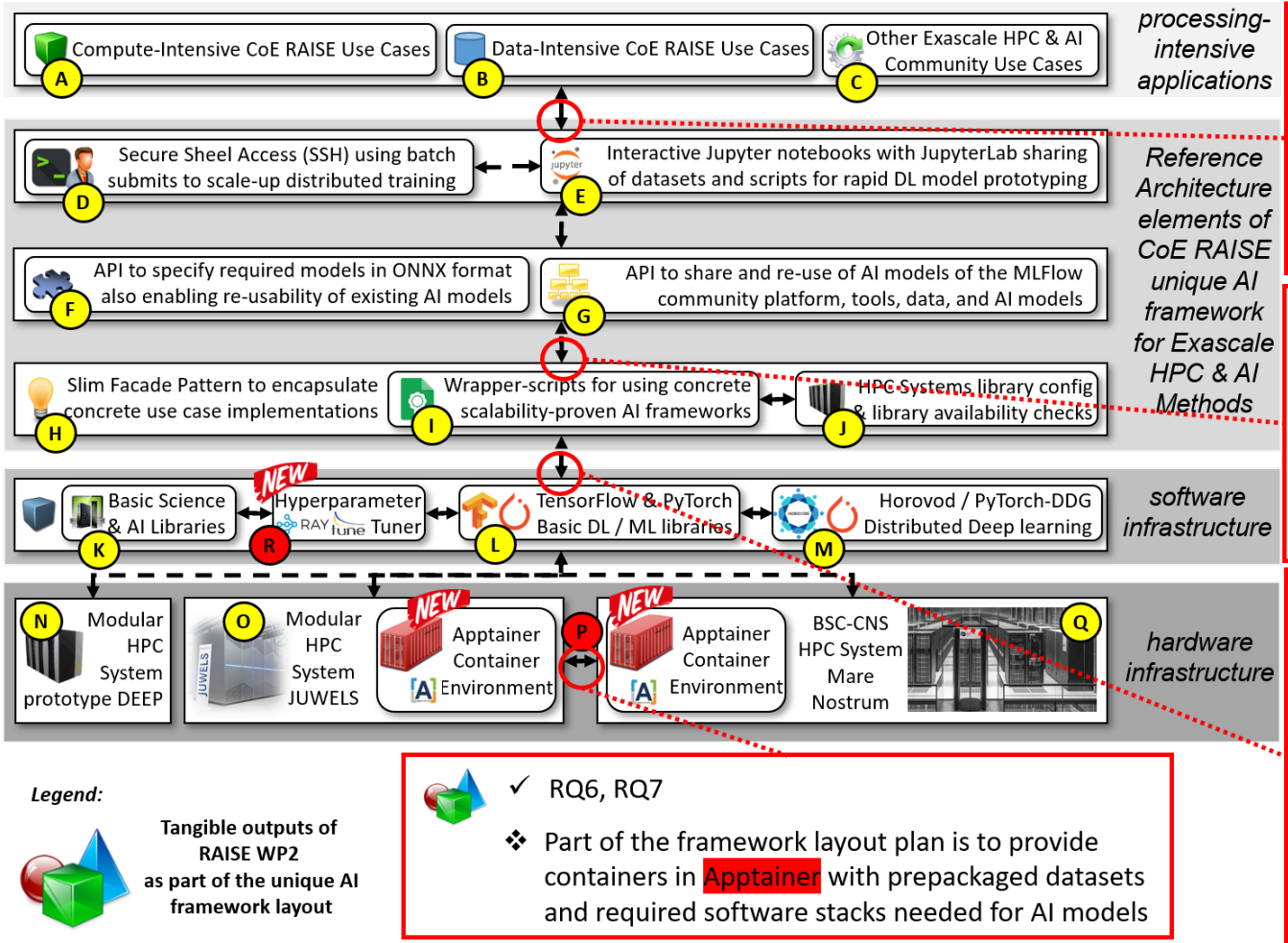
- CoR RAISE Interaction Room Process
  - Supports the proper software engineering design of the unique AI framework blueprint
  - Expecting to work with WP3 & WP4 experts in an open minded way
  - Process guided by Software Engineering Expert **Prof. Dr. Matthias Book** (University of Iceland)
  - Supported by Software Engineering & testing expert **Prof. Dr. Helmut Neukirchen** (University of Iceland)
  - CoE RAISE @ YouTube: https://www.youtube.com/channel/UCAdIZ-v6cWwGdapwYxdN7dg
  - Methology as one CoE RAISE outcome





Book, M., **Riedel, M.**, **Neukirchen, H.,** Goetz, M.: Facilitating Collaboration in High-Performance Computing Projects with an Interaction Room, in conference proceedings of the 4th ACM SIGPLAN International Workshop on Software Engineering for Parallel Systems (SEPS 2017), October 22-27, 2017, Vancouver, Canada

CoE RAISE Unique AI Software Framework for Exascale



✓ RQ1, RQ2, RQ4, RQ5

❖ Parts of the framework layout plan is to provide Kernels for Jupyter notebooks with correct version setups of modules for specific HPC Systems

✓ RQ3, RQ6 NEW

❖ Parts of the framework layout plan is to provide lightweight & abstract Python APIs building on ONNX enabling exchange with MLFlow, OpenML, ClearML, etc.

✓ RQ1, RQ2

❖ Parts of the framework layout plan is to provide a lightweight Python API that abstracts from low level versioning of AI packages (with proven scalability) and is harmonized with different available HPC system module versions

✓ RQ6, RQ7

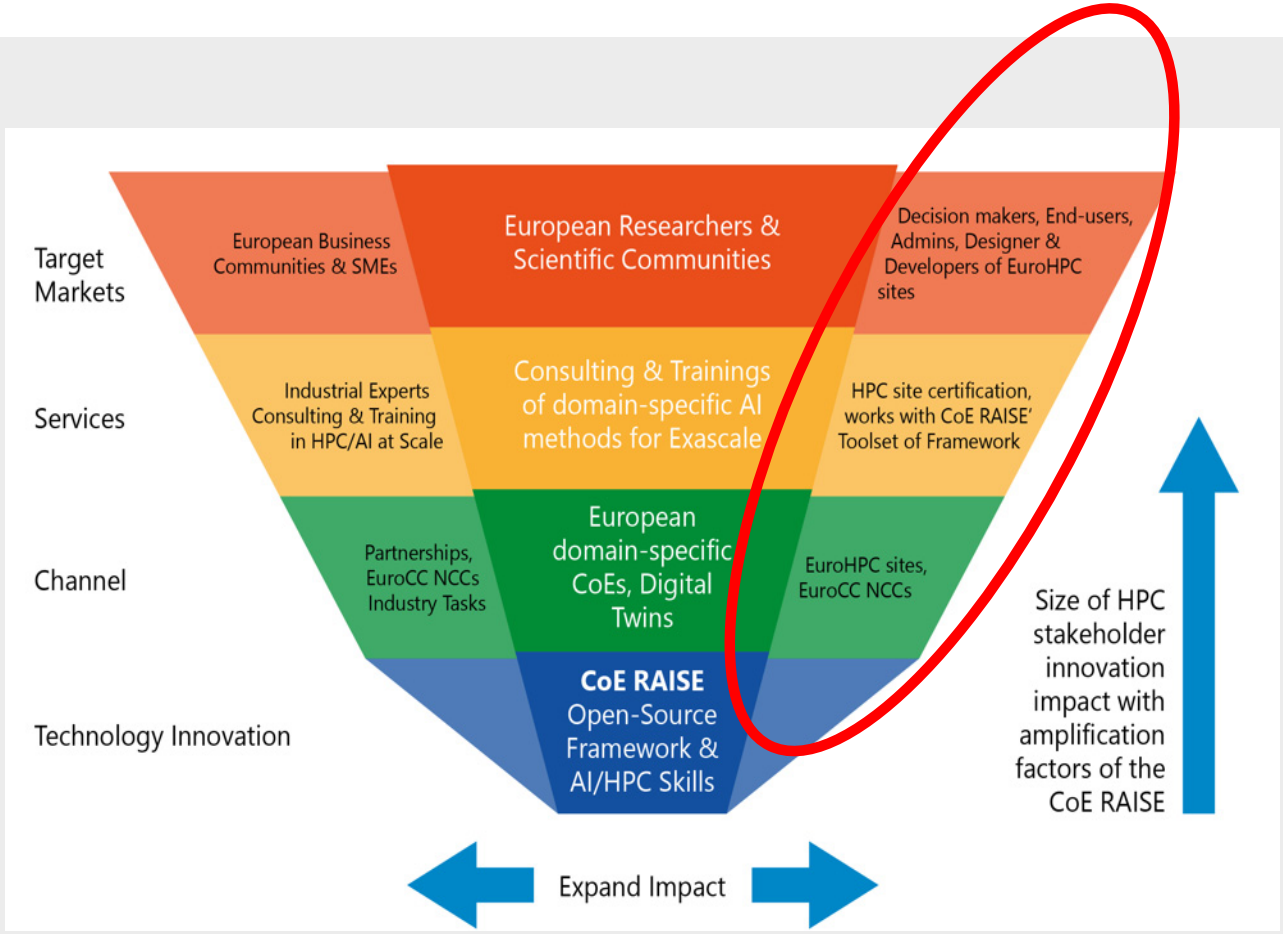
❖ Part of the framework layout plan is to provide containers in Apptainer with prepackaged datasets and required software stacks needed for AI models

Legend:

Tangible outputs of RAISE WP2 as part of the unique AI framework layout

Continously Updating

Towards SW Framework Adoptions



Work in Progress

LUMI

LUMI is a pre-exascale EuroHPC supercomputer located in Kajaani, Finland. It is a Cray EX supercomputer supplied by Hewlett Packard Enterprise (HPE) and hosted by CSC – IT Center for Science.

LUMI supercomputer
CSC

375 petaflops
Sustained performance

LEONARDO

Leonardo is a pre-exascale EuroHPC supercomputer currently built in the Bologna Technopole, Italy. It is supplied by ATOS, based on a BullSequana XH2000 supercomputer and hosted by CINECA.

LEONARDO Supercomputer
Cineca

249,47 petaflops
Sustained performance

323,40 petaflops
Peak performance

MARENOSTRUM 5

MareNostrum 5 is a pre-exascale EuroHPC supercomputer to be located in Barcelona, Spain. The system is supplied by Bull SAS combining Bull Sequana XH3000 and Lenovo ThinkSystem architectures. MareNostrum 5 is hosted by Barcelona Supercomputing Center (BSC).

New BSC's data centre waiting to host MN5 supercomputer
BSC

205 Petaflops
Sustained performance

314 Petaflops
Peak performance

VEGA

Vega is a petascale EuroHPC supercomputer located in Maribor, Slovenia. It is supplied by Atos, based on the BullSequana XH2000 supercomputer and hosted by IZUM.

VEGA supercomputer
IZUM

6,92 petaflo… …05 petaflops
Sustained perfor… … Peak …

MELUXINA

MeluXina is a petascale EuroHPC supercomputer located in Bissen, Luxembourg. It is supplied by Atos, based on the BullSequana XH2000 supercomputer platform and hosted by LuxProvide.

Meluxina supercomputer
LuxProvide

12,81petaflops
Sustained performance

18,29 petaflops
Peak performance

KAROLINA

Karolina is a petascale EuroHPC supercomputer located in Ostrava, Czech Republic. It is supplied by Hewlett Packard Enterprise (HPE), based on an HPE Apollo 2000Gen10 Plus and HPE Apollo 6500 supercomputers. Karolina is hosted by IT4Innovations National Supercomputing Center.

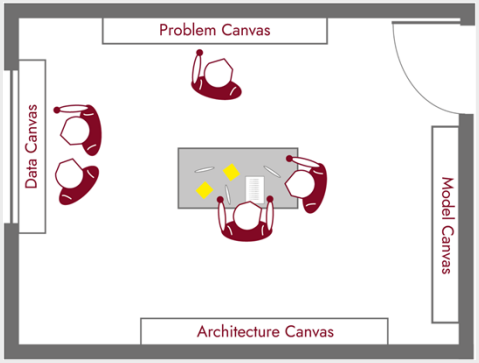
Karolina supercomputer
IT4Innovations

9,59 petaflops
Sustained performance

12,91 petaflops
Peak performance

Cross HPC/AI Methods Table – Transformer Models Interest

- ✓ Interaction Rooms
  - ✓ Update of Matrix
  - ✓ Components relatively constant & common
  - ✓ Methods change & new methods added (e.g., Transformer Models)



Use Case	AE	PINN	ANNs		CNN		NO	GNN		RNN		GAN	TF				SVM	RF
Details	*CAE*		*ANN*	*RBF-ANN*	*U-Net*	*RES NET*	*FNO*	*MLPF*	*GAT*	*LSTM*	*GRU*	*WGAN*	*MViT*	*ViViT*	*Swin*			
AI for turbulent boundary layers	X	X	X									X						
AI for wind farm layout optimization				X												X		
AI for data-driven models in reacting flows					X				X									
Smart models for next generation aircraft engine design					X				X									
AI for wetting hydrodynamics	X	X					X			X								
Event reconstruction and classification at the CERN HL-LHC use case								X										
Seismic imaging with remote sensing for energy applications	X	X				X	X			X	X					X	X	X
Detect-free metal additive manufacturing	X		X									X	X	X	X			
Sound Engineering	X		X															

drive. enable. innovate.

RAISE Center of Excellence

The CoE RAISE project receives funding from
the European Union's Horizon 2020 –
Research and Innovation Framework Programme
H2020-INFRAEDI-2019-1 under grant agreement no. 951733