



EuroHPC
Joint Undertaking



T2.4 - Software design of a unique AI framework Status & next steps – Extension Period Discussions (**M42+**)

Prof. Dr. – Ing. Morris Riedel et al.

Full Professor, School of Engineering & Natural Sciences, University of Iceland

Lead of CoE RAISE WP2 – AI- and HPC-Cross Methods at Exascale

2023-08-28, AHM RAISE, Hveragerði, Iceland



@ProfDrMorrisRiedel



@Morris Riedel



@MorrisRiedel



@MorrisRiedel



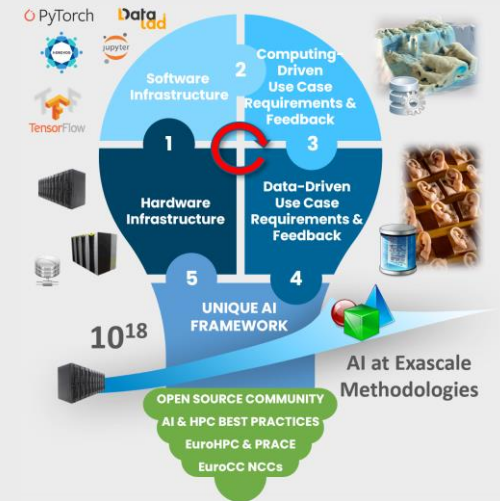
<https://www.youtube.com/channel/UCWC4VKHmL4NZgFfKoHtANKg>



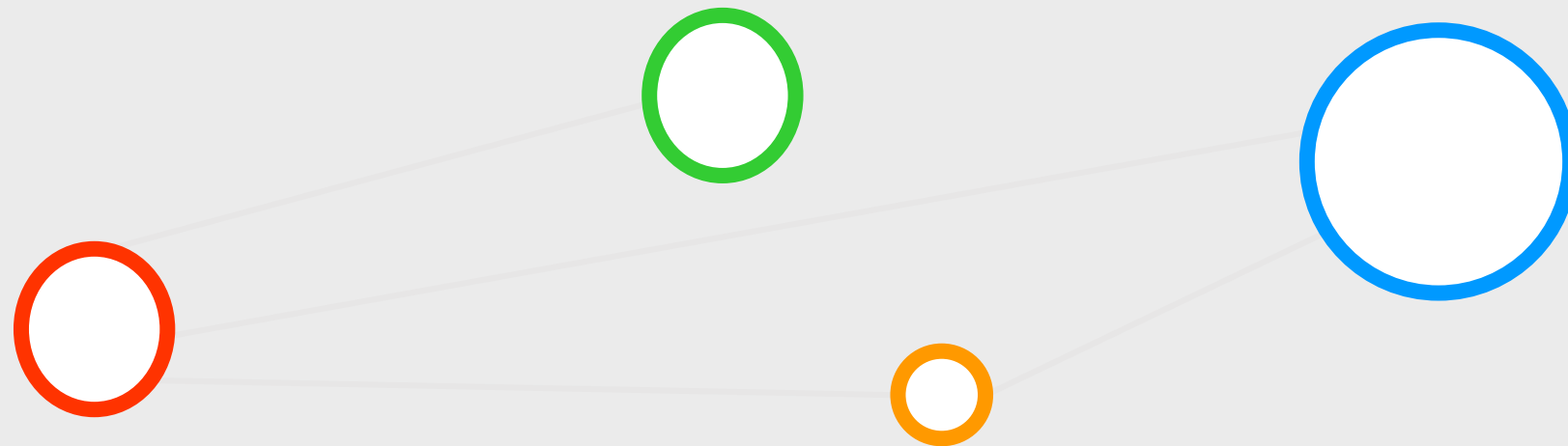
morris@hi.is

Outline

- Task 2.4 Process
 - TrustLLM as a new UAIF Contribution
- Challenges in using AI Methods on HPC at Scale
 - Review Toolset & Skillset Challenges
- Unique AI Framework (UAIF) Co-Design Process
 - UAIF Co-Design at A Glance
 - Factsheets & Interaction Rooms
- CoE RAISE UAIF Status
 - Current Blueprint
- Adoption Roadmap of the Framework
 - Cooperation with NCCs & EuroHPC JU Hosting Sites
- Summary & Q&A
 - Feedback from NCCs, CoEs



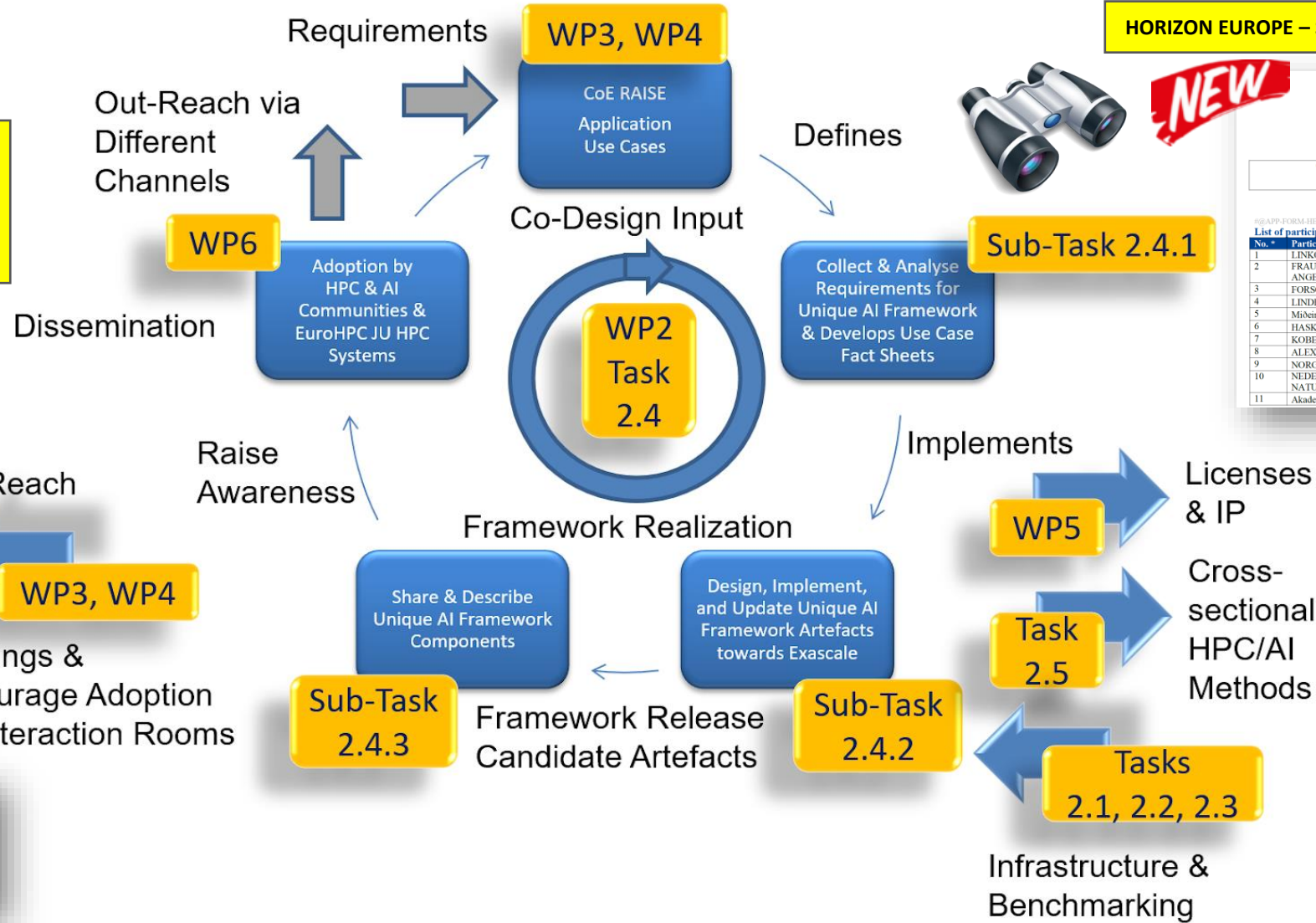
T2.4 Process



Task 2.4 Process – Focus on Adoption & New Stakeholders



Extension Period Discussions: Project members said: „we can demonstrate towards the end of the project, but not now“ – what training we can do towards the end now really together?



HORIZON EUROPE – START 11/2023 – 3 years – 14.5/15 Score



TRUSTLLM – DEMOCRATIZING TRUSTWORTHY AND EFFICIENT LARGE LANGUAGE MODEL TECHNOLOGY FOR EUROPE

#APP-FORM-HERIAA@#
List of participants

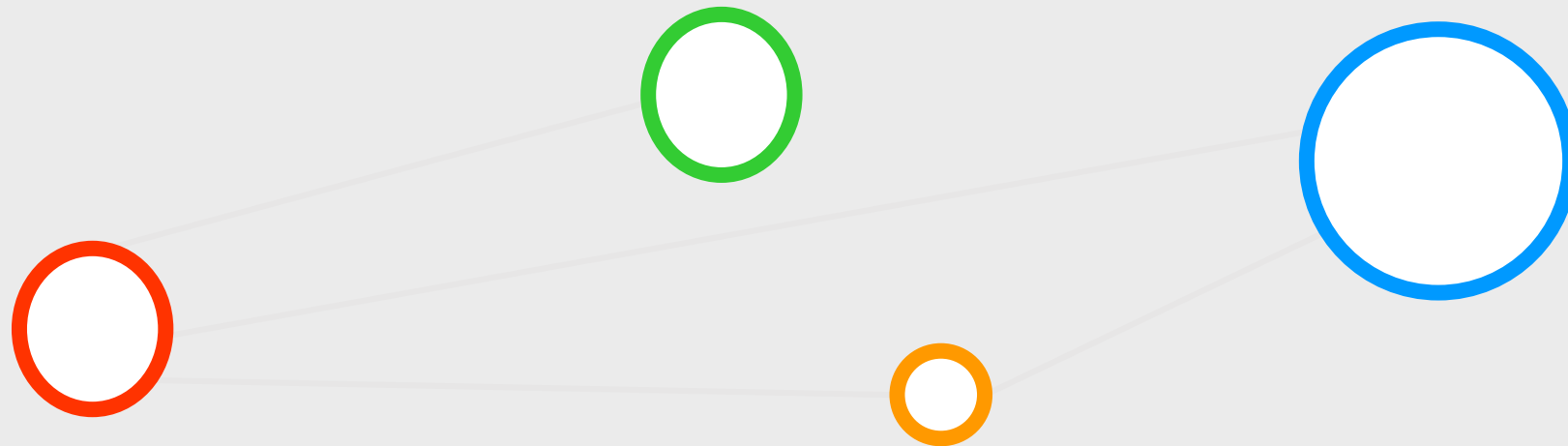
No. #	Participant organisation name	Acronym	Country
1	LINKÖPINGS UNIVERSITET (Coordinator)	LIU	SE
2	FRÄUNHOFER GESELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG EV	FGH	DE
3	FORSCHUNGSZENTRUM JULICH GMBH	FZJ	DE
4	LINDHOLMEN SCIENCE PARK AKTIEBOLAG	LSP	SE
5	Mibend ehf.	MID	IS
6	HASKOLI ISLANDS	UOI	IS
7	KOBENHAVNS UNIVERSITET	UCPH	DK
8	ALEXANDRA INSTITUTTET A/S	AXI	DK
9	NORGES TEKNISKE-NATURVITENSKAPELIGE UNIVERSITET	NTNU	NO
10	NEDERLANDSE ORGANISATIE VOOR TOEGEPAST NATUURWETENSCHAPPELIJK ONDERZOEK	TNO	NL
11	Akademie für Künstliche Intelligenz AKI gGmbH	AKI	DE

Extension Period Discussions: New AI areas, new HPC systems, new use cases for UAIF, new UAIF components?

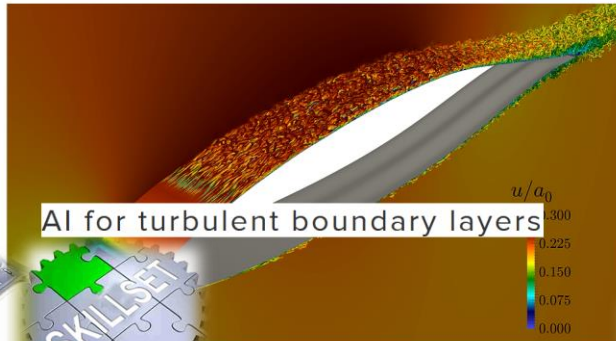
Extension Period Discussions: One example of new AI areas are Large Language Models (LLMs) & foundational models – Synergies beyond JUELICH/FZJ & UOI: Kurt (FM), Others???



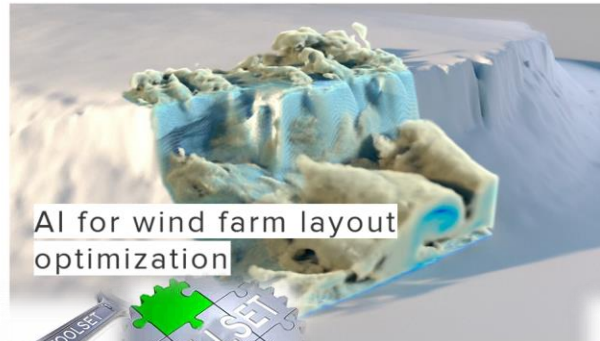
Challenges in using AI Methods on HPC at Scale



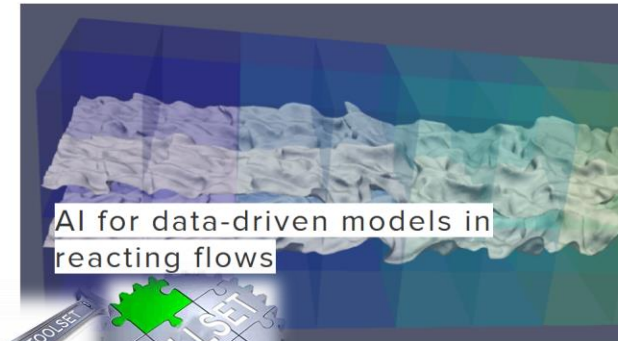
Compute & Data-driven Use Cases – Complex Challenges



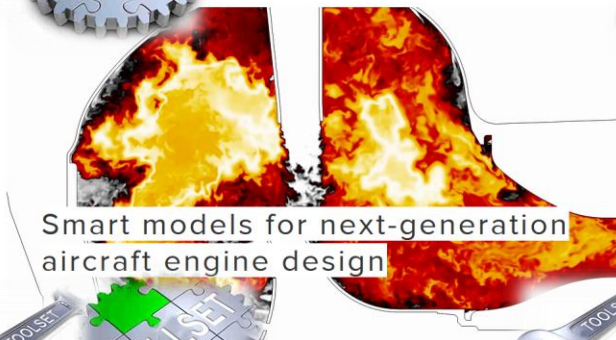
AI for turbulent boundary layers



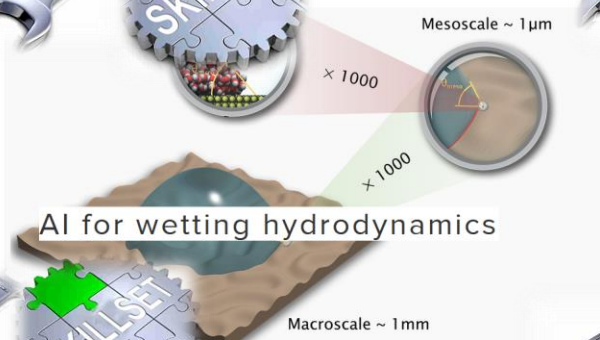
AI for wind farm layout optimization



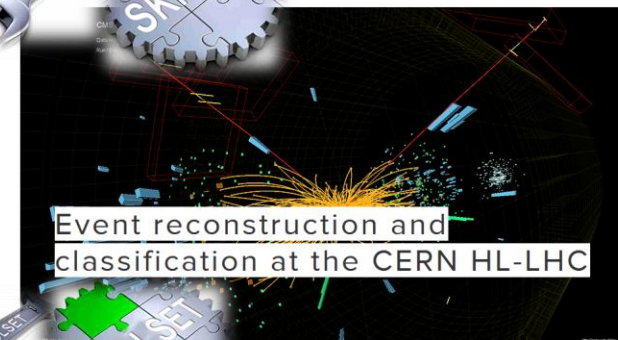
AI for data-driven models in reacting flows



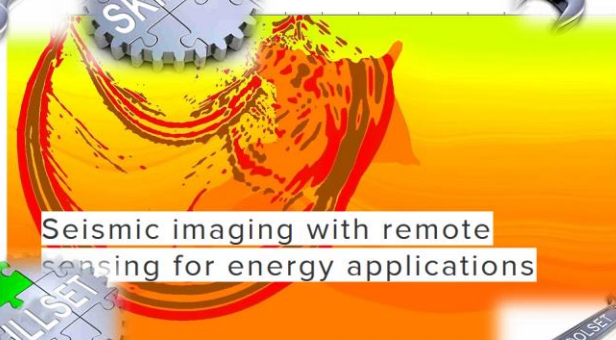
Smart models for next-generation aircraft engine design



AI for wetting hydrodynamics



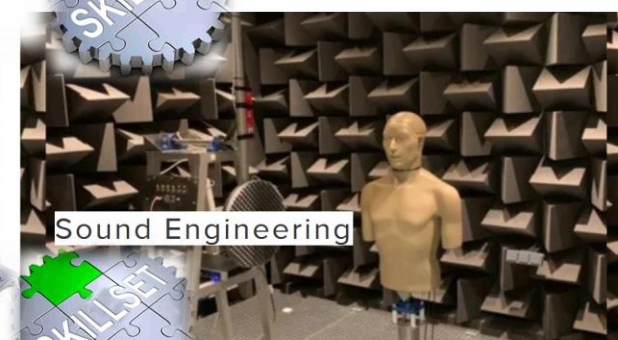
Event reconstruction and classification at the CERN HL-LHC



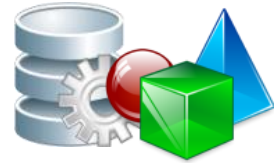
Seismic imaging with remote sensing for energy applications



Defect-free metal additive manufacturing



Sound Engineering

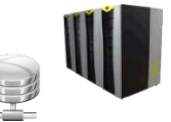


PyTorch



TensorFlow

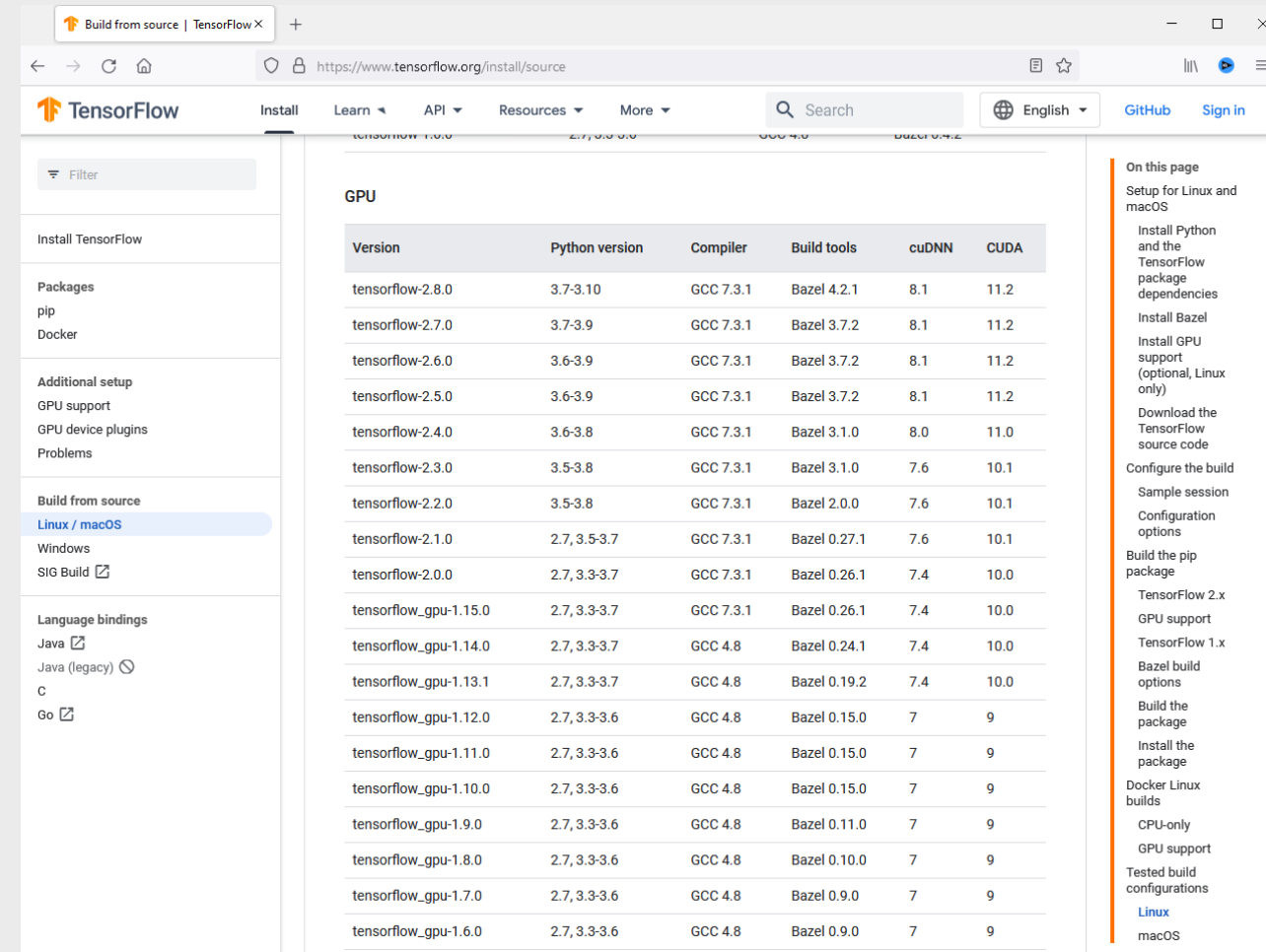
DataLad



Requirements Gathering Process – Version Challenges

➤ Example: TensorFlow

- Can we create an automated module checker for the SW Framework RAISE?
- Specific versions of TensorFlow require specific versions of underlying HPC modules or other AI frameworks to work in specific versions together
- Python versions must be correct as well
- E.g., differences in Python 3.8.x and 3.9.x
- **Support AI developers for many other tools like PyTorch, Horovod, Ray Tune, etc.**



The screenshot shows the TensorFlow installation page for Linux/macOS. The main content is a table titled "GPU" with columns for Version, Python version, Compiler, Build tools, cuDNN, and CUDA. The table lists various TensorFlow versions and their corresponding dependencies.

Version	Python version	Compiler	Build tools	cuDNN	CUDA
tensorflow-2.8.0	3.7-3.10	GCC 7.3.1	Bazel 4.2.1	8.1	11.2
tensorflow-2.7.0	3.7-3.9	GCC 7.3.1	Bazel 3.7.2	8.1	11.2
tensorflow-2.6.0	3.6-3.9	GCC 7.3.1	Bazel 3.7.2	8.1	11.2
tensorflow-2.5.0	3.6-3.9	GCC 7.3.1	Bazel 3.7.2	8.1	11.2
tensorflow-2.4.0	3.6-3.8	GCC 7.3.1	Bazel 3.1.0	8.0	11.0
tensorflow-2.3.0	3.5-3.8	GCC 7.3.1	Bazel 3.1.0	7.6	10.1
tensorflow-2.2.0	3.5-3.8	GCC 7.3.1	Bazel 2.0.0	7.6	10.1
tensorflow-2.1.0	2.7, 3.5-3.7	GCC 7.3.1	Bazel 0.27.1	7.6	10.1
tensorflow-2.0.0	2.7, 3.3-3.7	GCC 7.3.1	Bazel 0.26.1	7.4	10.0
tensorflow_gpu-1.15.0	2.7, 3.3-3.7	GCC 7.3.1	Bazel 0.26.1	7.4	10.0
tensorflow_gpu-1.14.0	2.7, 3.3-3.7	GCC 4.8	Bazel 0.24.1	7.4	10.0
tensorflow_gpu-1.13.1	2.7, 3.3-3.7	GCC 4.8	Bazel 0.19.2	7.4	10.0
tensorflow_gpu-1.12.0	2.7, 3.3-3.6	GCC 4.8	Bazel 0.15.0	7	9
tensorflow_gpu-1.11.0	2.7, 3.3-3.6	GCC 4.8	Bazel 0.15.0	7	9
tensorflow_gpu-1.10.0	2.7, 3.3-3.6	GCC 4.8	Bazel 0.15.0	7	9
tensorflow_gpu-1.9.0	2.7, 3.3-3.6	GCC 4.8	Bazel 0.11.0	7	9
tensorflow_gpu-1.8.0	2.7, 3.3-3.6	GCC 4.8	Bazel 0.10.0	7	9
tensorflow_gpu-1.7.0	2.7, 3.3-3.6	GCC 4.8	Bazel 0.9.0	7	9
tensorflow_gpu-1.6.0	2.7, 3.3-3.6	GCC 4.8	Bazel 0.9.0	7	9

Requirements Gathering Process – Module Challenges

➤ Example of Setups

- Many different versions / combinations
- E.g. FZJ JSC DEEP-EST HPC System

```
[riedell@dp-dam01 ~]$ module spider nccl
-----
NCCL:
-----
Description:
The NVIDIA Collective Communications Library (NCCL) implements multi-GPU and multi-node collective communication primitives that are performance optimized for NVIDIA GPUs.

Versions:
NCCL/2.4.2-1-CUDA-9.2.88
NCCL/2.4.6-1-CUDA-10.1.105
NCCL/2.4.8-CUDA-10.1.105
NCCL/2.4.8
NCCL/2.7.3-1-CUDA-10.2.89
NCCL/2.8.3-1-CUDA-11.0
NCCL/2.8.3-1-CUDA-11.3
NCCL/2.10.3-1-CUDA-11.3
NCCL/2.11.4-CUDA-11.5

For detailed information about a specific "NCCL" module (including how to load the modules) use the module's full name.
For example:
$ module spider NCCL/2.7.3-1-CUDA-10.2.89
-----
```

```
[riedell@dp-dam01 ~]$ module spider cuda
-----
CUDA:
-----
Description:
CUDA (formerly Compute Unified Device Architecture) is a parallel computing platform and programming model created by NVIDIA and implemented by the graphics processing units (GPUs) that they produce. CUDA gives developers access to the virtual instruction set and memory of the parallel computational elements in CUDA GPUs.

Versions:
CUDA/9.2.88
CUDA/10.1.105
CUDA/10.2.89
CUDA/11.0
CUDA/11.0.207
CUDA/11.3
CUDA/11.5

For detailed information about a specific "CUDA" module (including how to load the modules) use the module's full name.
For example:
$ module spider CUDA/11.0.207
-----
```

```
[riedell@dp-dam01 ~]$ module spider cudnn
-----
cudNN:
-----
Description:
The NVIDIA CUDA Deep Neural Network library (cuDNN) is a GPU-accelerated library of primitives for deep neural networks.

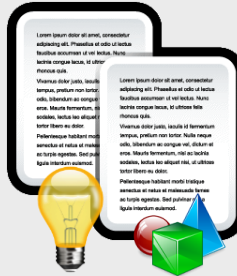
Versions:
cudNN/7.4.1.5-CUDA-9.2.88
cudNN/7.5.1.10-CUDA-10.1.105
cudNN/7.6.4.38-CUDA-10.1.105
cudNN/7.6.5.32-CUDA-10.2.89
cudNN/8.0.2.39-CUDA-11.0
cudNN/8.2.1.32-CUDA-11.3
cudNN/8.3.1.22-CUDA-11.5

For detailed information about a specific "cuDNN" module (including how to load the modules) use the module's full name.
For example:
$ module spider cudNN/7.6.5.32-CUDA-10.2.89
-----
```

```
[riedell@dp-dam01 ~]$ module spider tensorflow
-----
TensorFlow:
-----
Description:
An open-source software library for Machine Intelligence

Versions:
TensorFlow/1.12.0-GPU-Python-2.7.15
TensorFlow/1.12.0-GPU-Python-3.6.6
TensorFlow/1.13.1-GPU-Python-3.6.8
TensorFlow/2.2.0-GPU-Python-3.6.8-1
TensorFlow/2.3.1-Python-3.8.5
TensorFlow/2.5.0-Python-3.8.5
TensorFlow/2.6.0-CUDA-11.5

For detailed information about a specific "TensorFlow" module (including how to load the modules) use the module's full name.
For example:
$ module spider TensorFlow/2.2.0-GPU-Python-3.6.8-1
-----
```



Example: Detailed Knowledge of Modules Necessary

➤ Modules

- Vary heavily between different HPC systems
- 2-3 Days/Months spend by researchers for getting the right environment / HPC system
- Goal: UAIF simplify setup of components
- E.g., automated job script generator for right module setup
- E.g., re-usable scripts

```
#!/usr/bin/env bash

# Slurm job configuration
#SBATCH --nodes=1
#SBATCH --ntasks-per-node=4
#SBATCH --cpus-per-gpu=20
#SBATCH --account=hai_so2sat
#SBATCH --output=output.out
#SBATCH --error=error.er
#SBATCH --time=6:00:00
#SBATCH --job-name=BENTF2
#SBATCH --gres=gpu:1 --partition=booster

#load modules
ml Stages/2020 GCC/9.3.0 OpenMPI/4.1.0rc1
ml Horovod/0.20.3-Python-3.8.5
ml TensorFlow/2.3.1-Python-3.8.5
#activate my virtualenv
#source /p/project/joaiml/remote_sensing/rocco_sedona/ben_TF2/scripts/env_tf2_juwels_booster/bin/activate

#export relevant env variables
#export CUDA_VISIBLE_DEVICES="0,1,2,3"

#run Python program
srun --cpu-bind=none python -u train_hvd_keras_aug.py
```

Deep_DDP	important bug fix	3 months ago
Deep_DeepSpeed	Deepspeed in Deep	6 months ago
Deep_HeAT	Jureca additions	5 months ago
Deep_Horovod	Deep modifications for Horovod and fex bu...	6 months ago
Deep_TensorFlow	initial TF push	5 months ago
HELPER_Scripts	fix tqdm bug	4 months ago
Jureca_DDP	latest fixes	1 month ago
Jureca_DeepSpeed	latest fixes	1 month ago
Jureca_Graphcore	added Graphcore dir and fixed lrank in CASES	2 months ago
Jureca_HeAT	latest fixes	1 month ago
Jureca_Horovod	latest fixes	1 month ago
Jureca_LibTorch	initial libtorch push	1 month ago
Jureca_RayTune	Update Jureca_RayTune/create_jureca_env.sh	3 months ago
Juwels_DDP	Update README.md	3 months ago
Juwels_Turbulence	merge	9 months ago
PARAMETER_TUNING	Update PARAMETER_TUNING/Autoencoder/...	3 months ago

Already available for the community: <https://gitlab.jsc.fz-juelich.de/CoE-RAISE/FZJ/ai-for-hpc-oa>

Requirements Gathering Process – Time Efforts Challenges

➤ Example of Setups

- Tried many varieties of kernels
- **Developers /PIs / PhD Students loose ~3-4 hours average by trying new HPC machine just to get new modules right and/or setup kernels that work with modules**
- **Selected debug/solution tools not known always, e.g., nvidia-smi, really scalable components, etc.**
- **Note: Jupyter framework itself seems not to be the problem, rather complex hardware/software configurations**

The screenshot shows the 'Start Links' page of the Jülich Supercomputing Centre. It features a 'Configurations' section with a table listing four JupyterLab instances. Each instance has a 'Name', 'Version', 'System', 'Account', 'Project', 'Partition', 'Details', and 'Actions' column. The 'Actions' column contains 'Start' and 'Delete' buttons for each instance.

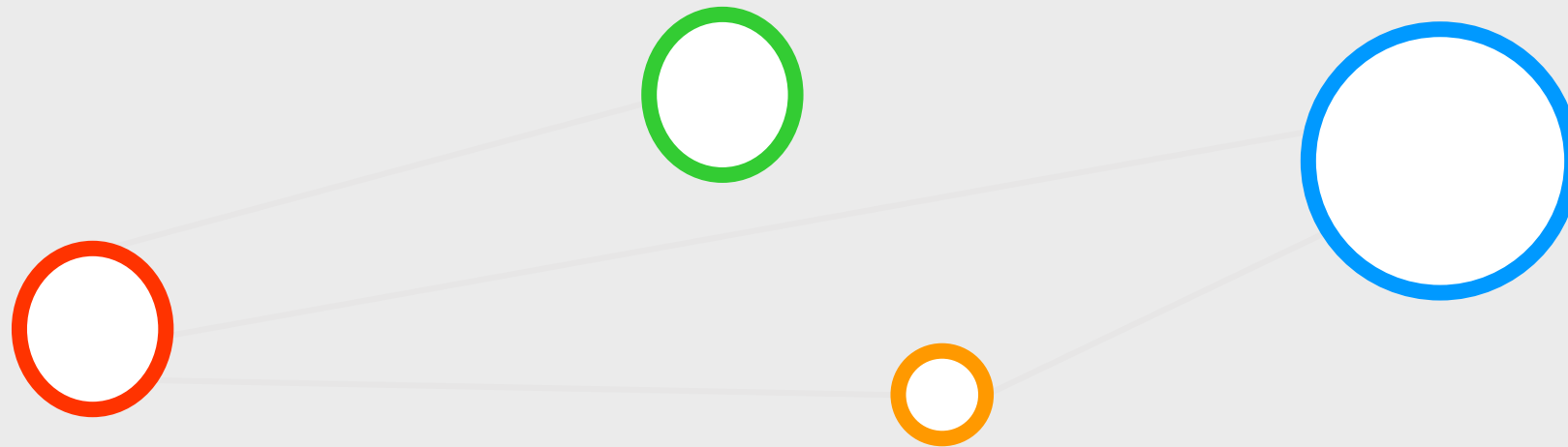
Name	Version	System	Account	Project	Partition	Details	Actions
jupyterlab_1	JupyterLab 2 (2020b)	DEEP	riedel1	joaiml	dp-dam	► Details	Start Delete
jupyterlab_2	JupyterLab 2 (2020b)	DEEP	riedel1	joaiml	ml-gpu	► Details	Start Delete
jupyterlab_3	JupyterLab 2 (2020b)	DEEP	riedel1	joaiml	ml-gpu	► Details	Start Delete
jupyterlab_4	JupyterLab 2 (2020b)	DEEP	riedel1	joaiml	ml-gpu	► Details	Start Delete

The screenshot shows a JupyterLab code editor with Python code for a neural network. A 'Select Kernel' dialog box is open, showing a list of available kernels. The 'Start Preferred Kernel' button is highlighted.

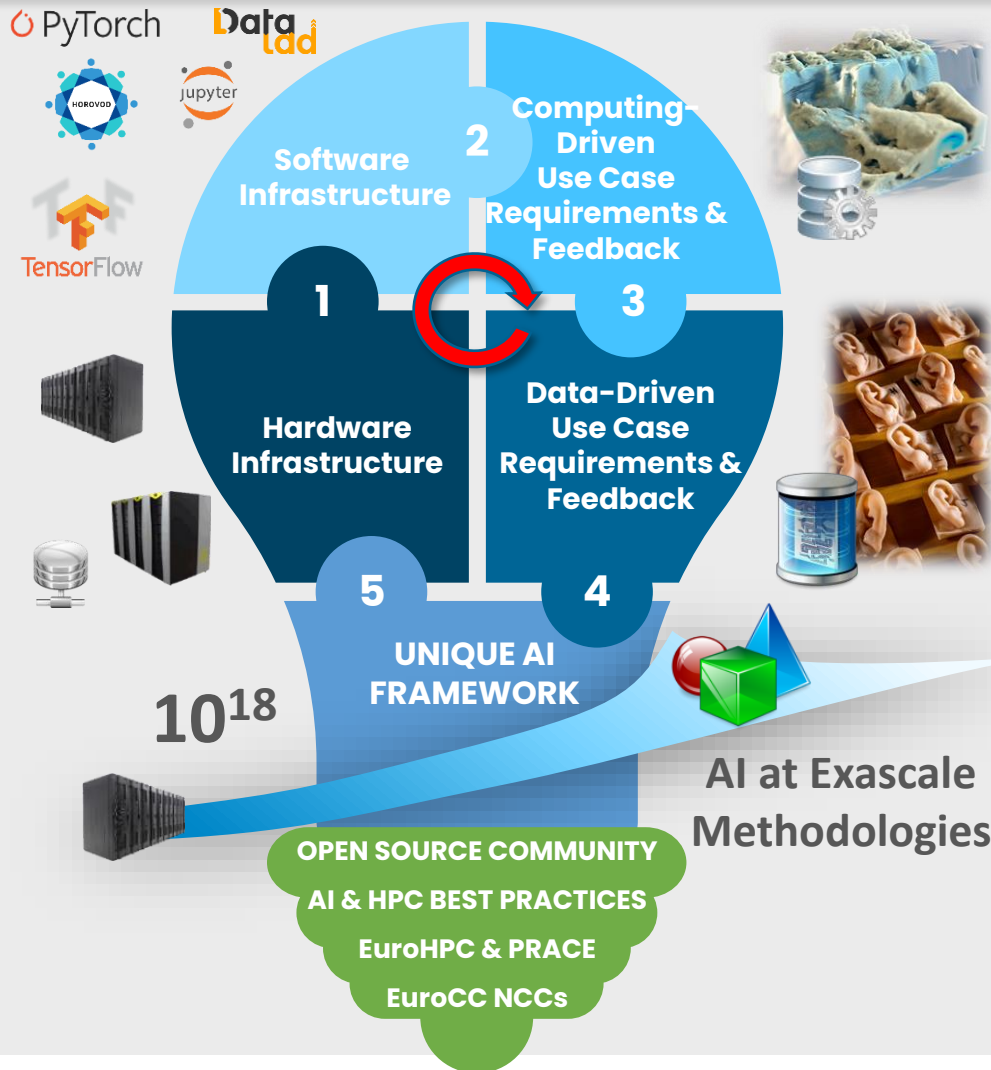
```
1 | import os
2 | os.environ['CUDA_VISIBLE_DEVICES'] = "0,1,2,3"
3 |
4 | # Device list
5 | device_list = os.environ['CUDA_VISIBLE_DEVICES'].split(',')
6 |
7 | # Load data
8 | (X_train, y_train), (X_test, y_test) = mnist.load_data()
9 |
10 | # Flatten 28x28 images to a 784 vector for each image
11 | num_pixels = X_train.shape[1] * X_train.shape[2]
12 | X_train = X_train.reshape((X_train.shape[0], num_pixels)).astype('float32')
13 | X_test = X_test.reshape((X_test.shape[0], num_pixels)).astype('float32')
14 | # Normalize inputs from 0-255 to 0-1
15 | X_train = X_train / 255
16 | X_test = X_test / 255
17 | # One-hot encode outputs
18 | y_train = to_categorical(y_train)
19 | y_test = to_categorical(y_test)
20 | num_classes = y_test.shape[1]
21 |
22 | # Define baseline model
23 | def baseline_model():
24 |     # create model
25 |     model = Sequential()
26 |     model.add(Dense(num_pixels, input_dim=num_pixels, kernel_initializer='normal', activation='relu'))
27 |     model.add(Dense(num_classes, kernel_initializer='normal', activation='softmax'))
28 |     # compile model
29 |     model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])
30 |     return model
31 |
32 | # Build the model
33 | model = baseline_model()
34 | # Fit the model
35 | model.fit(X_train, y_train, validation_data=(X_test, y_test), epochs=10, batch_size=200, verbose=2)
36 |
37 | # Final evaluation of the model
38 | scores = model.evaluate(X_test, y_test, verbose=0)
39 | print("Baseline Error: %.2f%% (%.2f score)" % (100 - scores[1], scores[1]))
40 |
```

Select Kernel
Select kernel for: 'CC_8D_ANN_Model.ipynb'
d_kernel
Start Preferred Kernel
d_kernel
d_kernel_students
d_kernel_students3
d_kernel_students4
d_kernel2
d_kernel3
d_kernel4
d_kernel5
d_kernel6
d_kernel7
d_kernel8
kernel_assignment
Octave-6.1.0
PyDeepLearning-1.0
PyTorch-1.8.1
PyQuantum-1.1
Python 3 (pykernel)

Unique AI Framework (UAIF) Co-Design Process



Unique AI Framework (UAIF) Co-Design Process at a Glance



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 (UAIF)

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

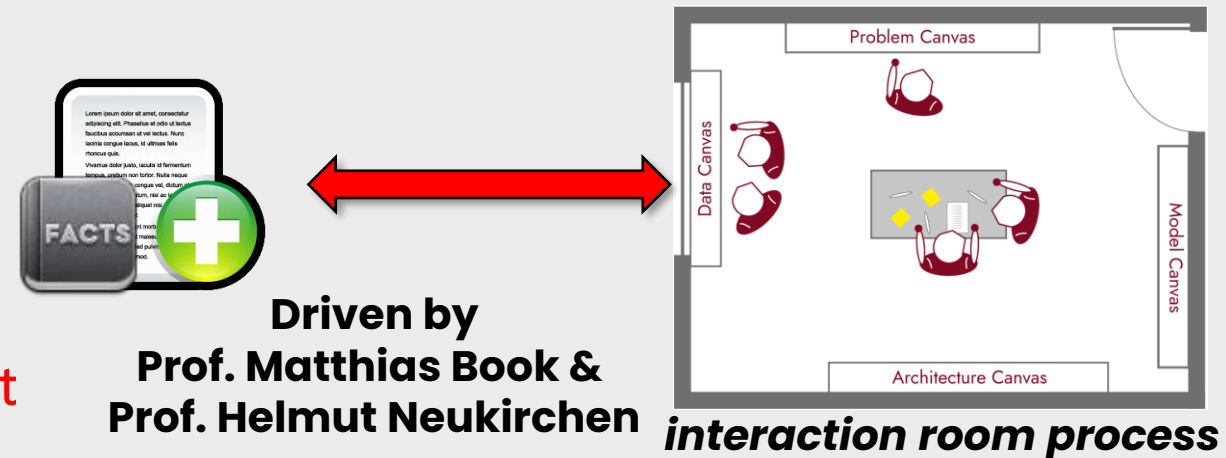
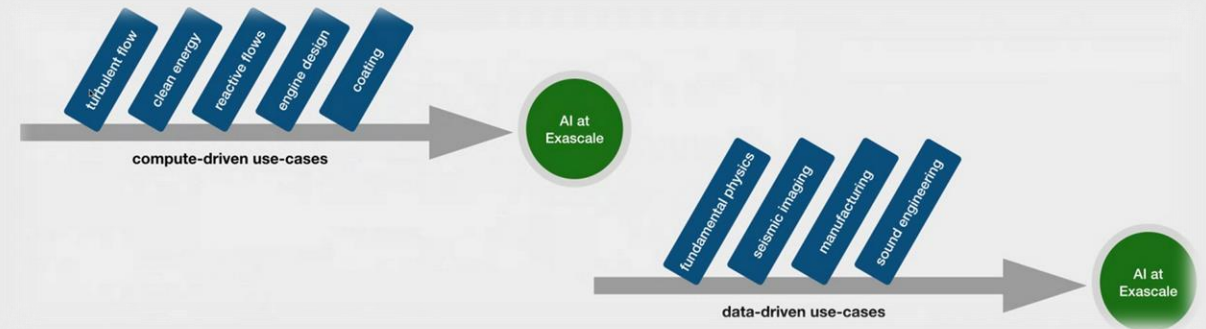
Unique AI Framework (UAIF) Co-Design Process Approaches

➤ 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**



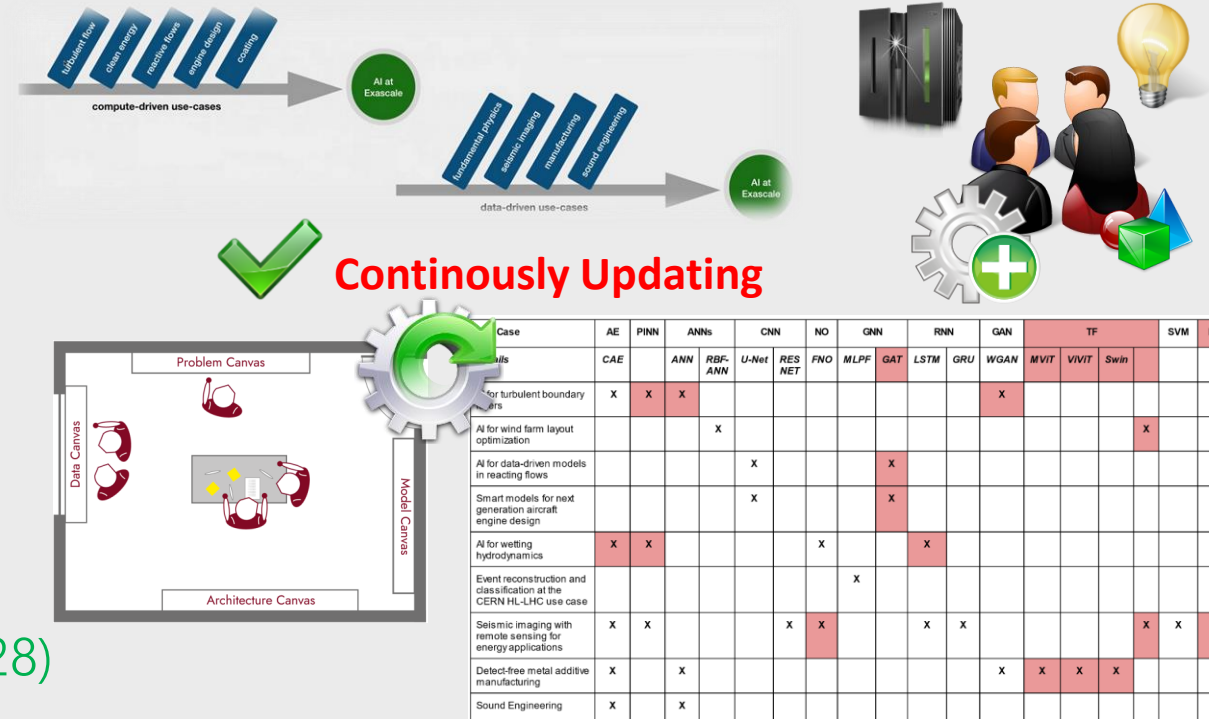
Interaction Room Status & Discussions – WP3/WP4 Overview

WP3 (third round IRs)

- T3.1: Turbulent Flow → Done (2023-06-05)
- T3.2: Clean Energy → Done (2023-04-11)
- T3.3: Reactive Flows → Done (2023-05-09)
- T3.4: Engine design → Done (2023-05-09)
- T3.5: Coating → Done (2023-04-24)

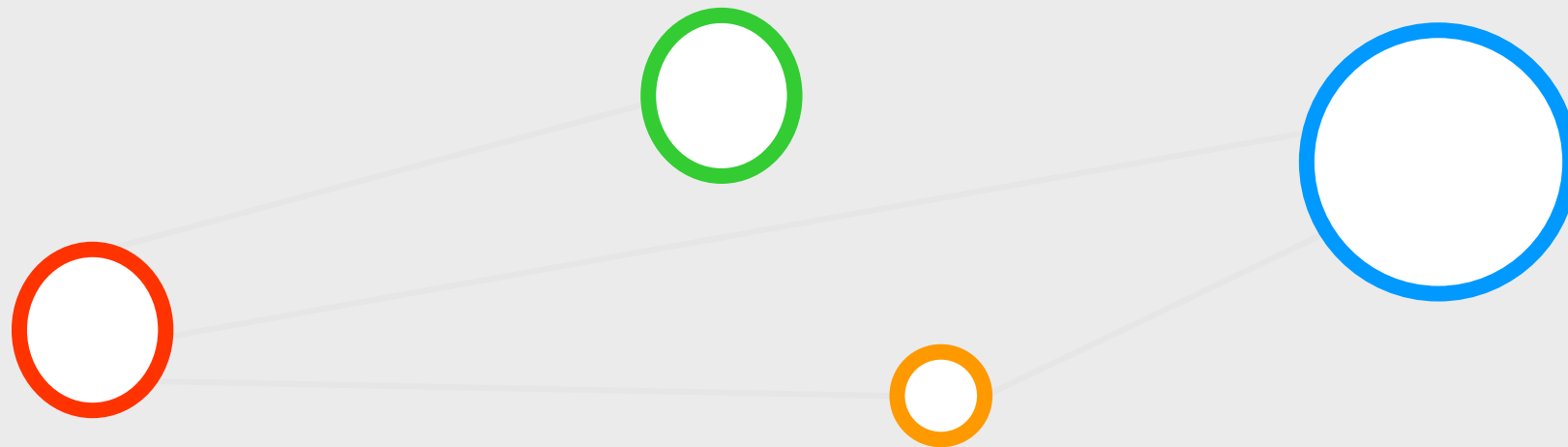
WP4 (third round IRs)

- T4.1: Fundamental physics → Done (2023-04-28)
- ⚠️ T4.2: Seismic imaging → September?
- T4.3: Manufacturing → Done (2023-05-02)
- T4.4: Sound engineering → Done (2023-04-21)
- 3rd iteration of Interaction Rooms → schedule

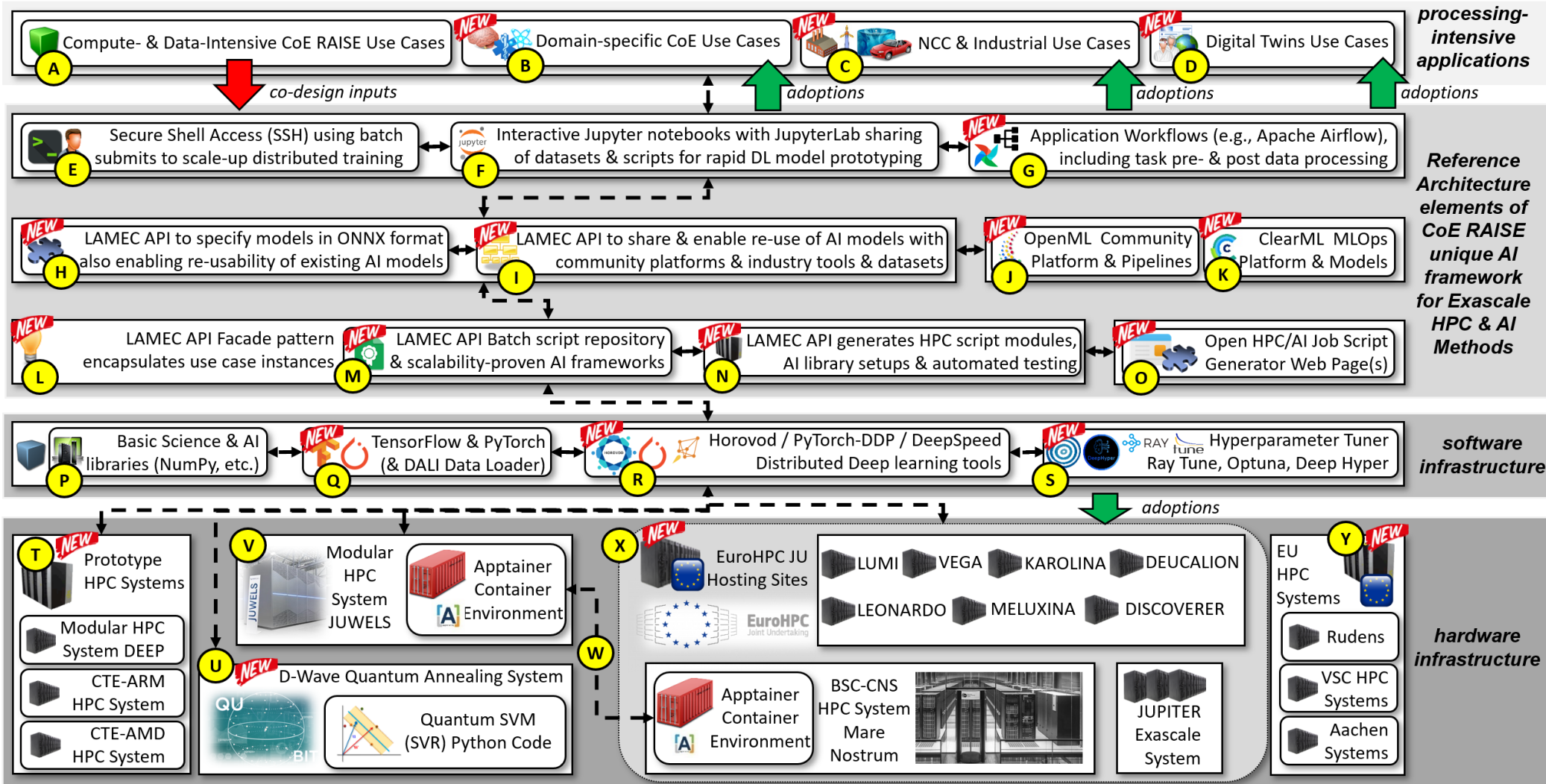


Next round Interaction Rooms for Adoption

- Carve out more details on AI/HPC methods
- Contribute to the Unique AI Framework
- Update our HPC/AI Methods Matrix



Realization of SW Framework



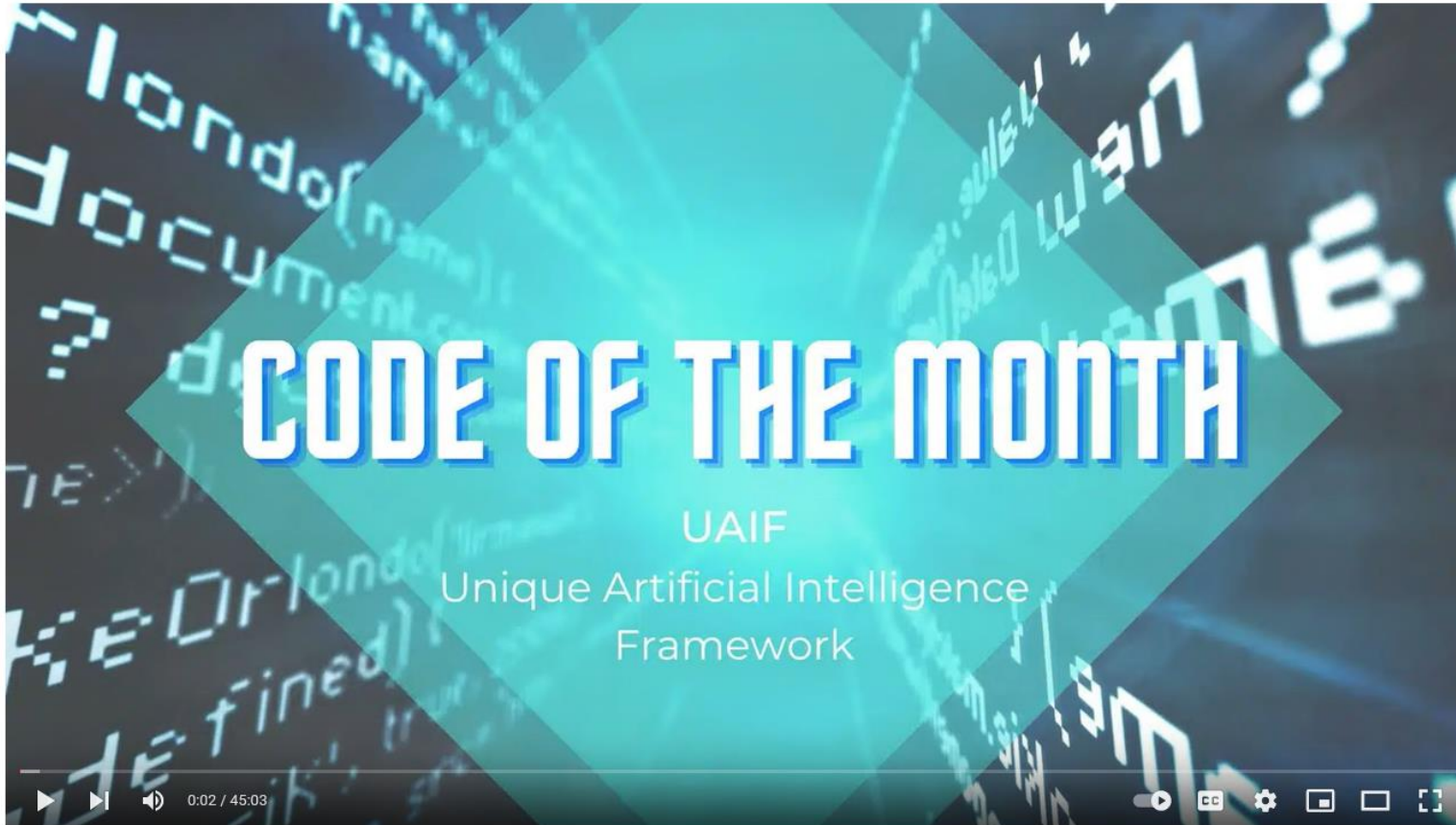
The strategy of “ready-to-run toolsets” Presented at CASTIEL Code of the Month event (2023-07-26 public Webinar), where to put the tools in the overview, e.g. HPC4AI, PhzDLL, etc.?

Continuously Updating!

Extension Period Discussions: LAMEC = Load AI Modules, Environments, and Containers – How far can we go? How many systems to add? What happens at M43? Sustainability? Calls?



Realization of SW Framework – YouTube RAISE Channel



Code of the month vol. 2 - Unique Artificial Intelligence Framework

RAISE CoE RAISE
77 subscribers

Subscribed

2 | Share | Download | Clip | Save

44 views 2 days ago

CoE RAISE follows the rules of open science and publishes its results open-access when they are ready for wider application. All developments of CoE RAISE are being integrated into the Unique AI Framework (UAIF), which will not only contain the trained models but also documentation on how to use them on current Petaflop and future Exascale HPC, prototype, and disruptive systems. The developments toward the Unique AI Framework are continuously progressing. [Show more](#)



Thanks to Lin, Eray,
Johannes, Arnar, ...
GREAT WORK!!!



**Continuously
Updating!**

Realization of SW Framework – Interactive Website

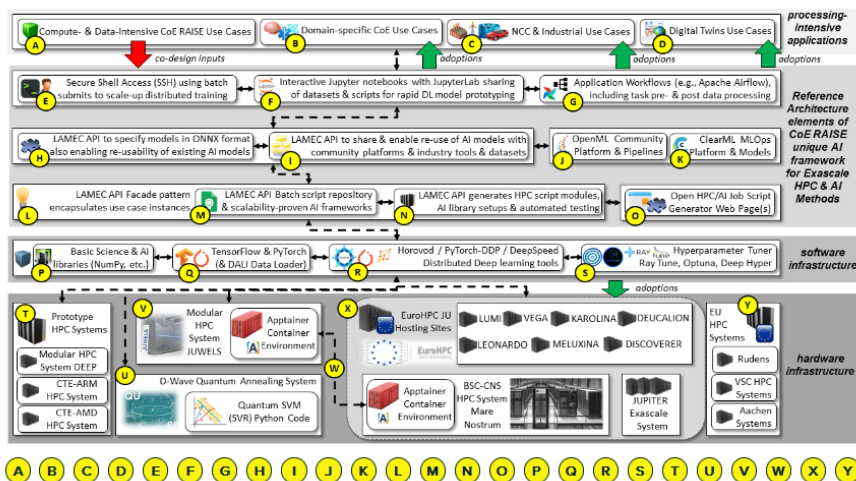


Unique AI Framework

- (UAIF) -

CoE RAISE follows the rules of open science and publishes its results open-access when they are ready for wider application. All developments of CoE RAISE are being integrated into the Unique AI Framework (UAIF), which will not only contain the trained models but also documentation on how to use them on current Petaflop and future Exascale HPC, prototype, and disruptive systems. The UAIF developed by CoE RAISE works with processing-intensive applications of a wide variety of scientific and engineering domains.

UAIF in the context of the larger European Ecosystem of Projects and Initiatives



A - Compute- and Data-Intensive CoE RAISE Use Cases

Component (A) in Fig. 1 represents the co-design efforts of the UAIF based on compute- and data-intensive use cases. Fact Sheets for each use case have been produced and describe what novel AI methods correlate to available UAIF components. They foster general understanding of the contributions that have been added over time to the UAIF and include scalability and utility for Exascale aspects. Several tasks in WP2 contributed to benchmarking and proof of scalability of selected components of the UAIF on various production and prototype HPC systems in this context. Detailed co-design activities have been performed via the Interaction Room methodology and Mural Boards. During the project and especially in the last reporting period, a clear picture is provided on what components are relevant for the UAIF.

B - Domain-Specific CoE Use Cases

A wide variety of CoEs have been funded in different domain-specific areas providing use cases that leverage simulation sciences or AI/HPC methods to utilize the emerging Exascale computing. At the time of writing, another EuroHPC JU Work Programme (WOPRO) outlining future funding of CoEs addresses the needs of large user communities in four specific areas of application domains. As shown in Fig. 1 (B), the UAIF is recommended to CoEs to adopt the UAIF to prevent AI developers in domain-specific sciences wasting a lot of effort.

C - NCC and Industrial Use Cases

A pan-European network of NCCs has been created under the EuroCC-1 and 2 to support Small and Medium Enterprises (SMEs) to leverage HPC resources made available via EuroHPC JU. The NCCs represent adoptions of the UAIF by NCCs and the significant potential to govern speed-up and scale-up their applications towards Exascale.

D - Digital Twins (DT) Use Cases

DTs and corresponding workflows as they are developed, e.g., in the Destination Digital Twins project, are highly relevant for scientific and engineering HPC users in Europe. Component (D) has been developed to support adoptions of the UAIF by DTs that are also highly relevant for CoE RAISE, either the DTs or the use cases in CoE RAISE.

Reference Architecture Elements

This section describes the reference architecture components relevant for the UAIF in the second layer (components E) – (O). This covers descriptions of the UAIF high-level access, application workflows, LAMEC API Open Neural Network API community platform integration, community platform OpenML interoperability, ClearML MLOps platform interoperability, LAMEC API facade pattern implementation, LAMEC API batch script repository, LAMEC API batch script generator, and open HPC/AI script generator web page(s).

E - Secure Shell (SSH) Low-Level Access

As shown in Fig. 1 (E), the first reference architecture element includes the use of the SSH protocol into the plan. Principally, as a means to remotely log into HPC systems and submit batch scheduler scripts via the Simple Linux Utility for Resource Management (SLURM) tool, it remains one of the integral access methods for HPC applications. It needs to be provided to researchers. One example of relevance for CoE RAISE is that AI researchers often use batch scripts for distributed training of DL models to leverage the high number of Graphical Processing Units (GPUs) that are available on HPC systems.

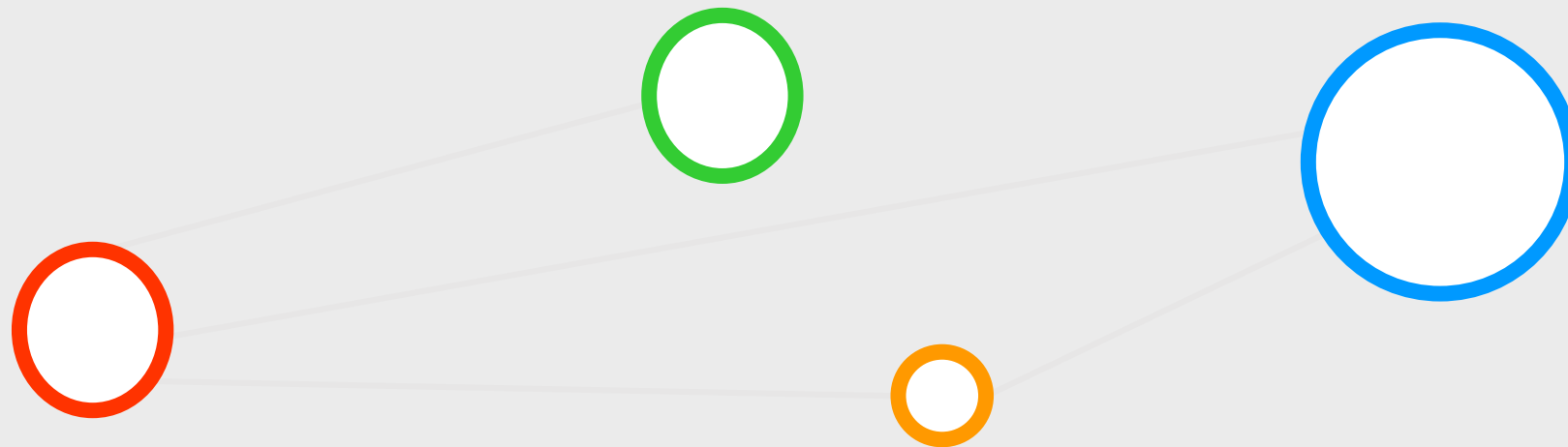


**Thanks to Michael & John, ...
GREAT WORK!!!**

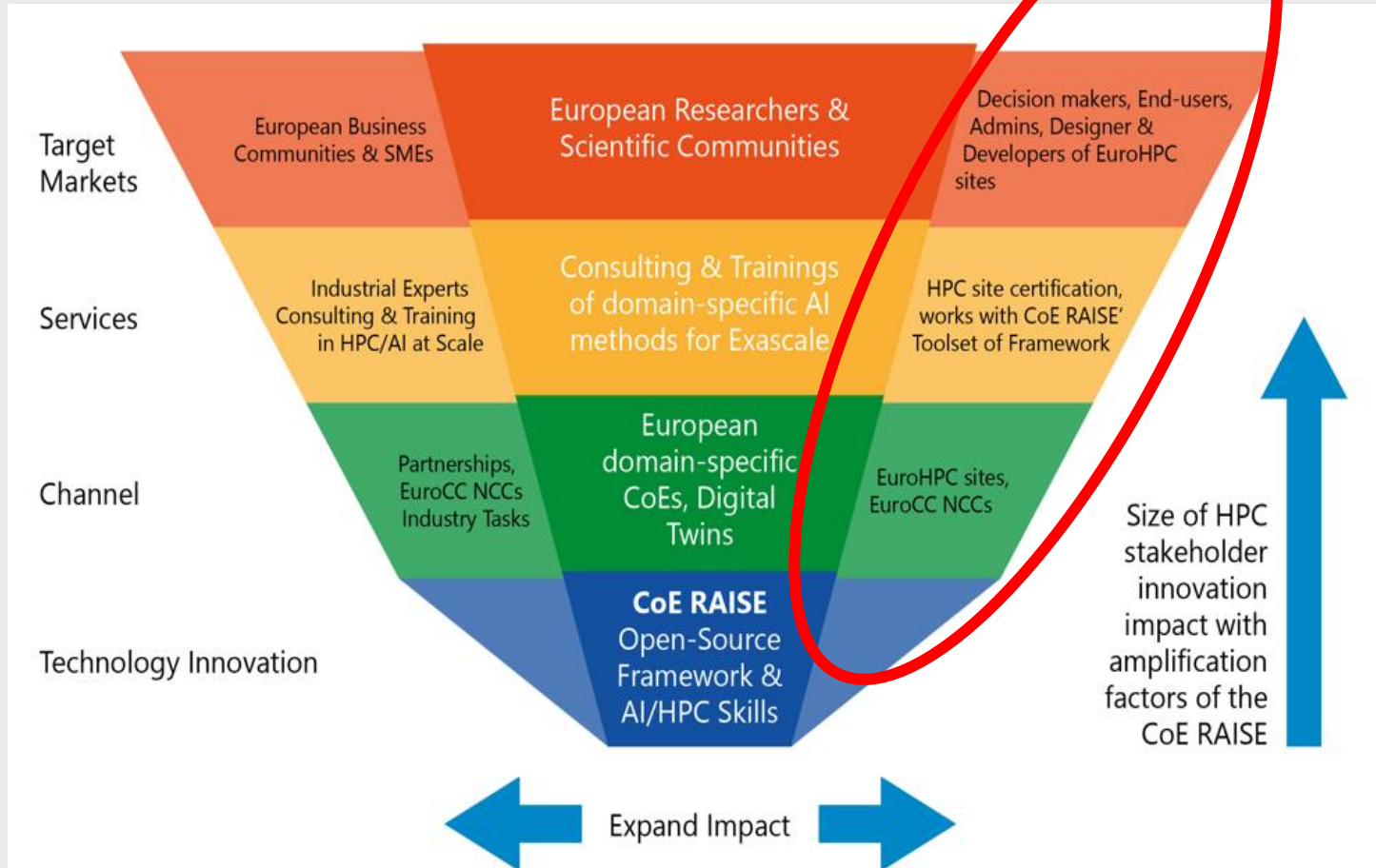


Continuously Updating!

Adoption Roadmap of the Framework



Adoption Roadmap of the Framework



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 **550 petaflops** Peak performance

LEONARDO

Leonardo is a pre-exascale EuroHPC supercomputer currently built in the Bologna Technopole. 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 XH2000 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 petaflops Sustained performance **10,05 petaflops** Peak performance

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

12,81 petaflops Sustained performance **18,29 petaflops** Peak performance

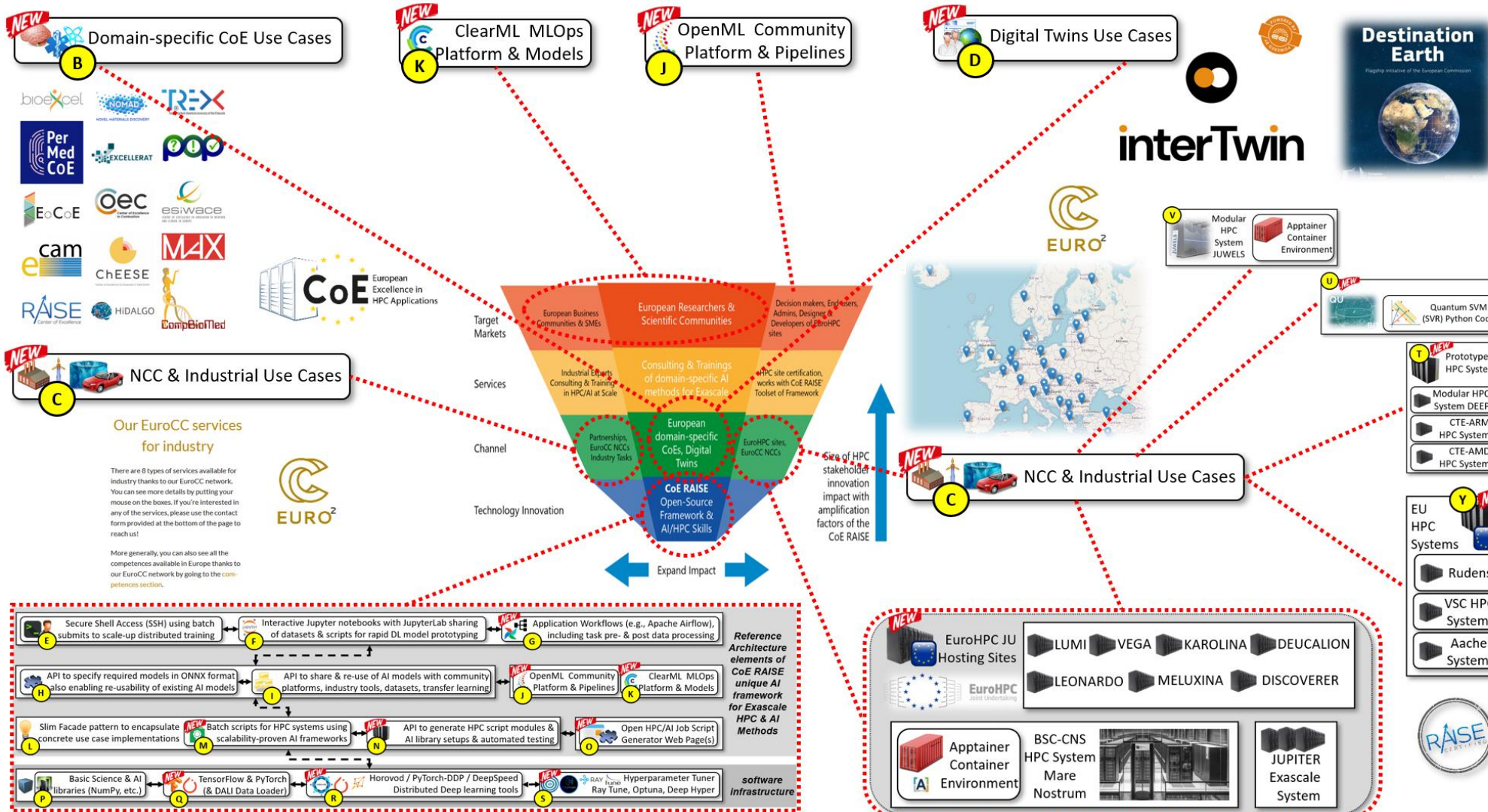
KAROLINA

Karolina is a petascale EuroHPC supercomputer located in Opatowitz, 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

Towards SW Framework Adoptions



NCCs & Industry Hosting Sites – Adoption Plans



Morris Riedel • You
 Professor & Head of Research Group High Productivity Data Processin...
 5d •

Productive meeting of EuroCC NCC Iceland/Croatia/Slovenia at IEEE MIPRO 2023 Conference crafting plans for #AI & #HPC Methods collaboration with CoE RAISE and its AI/HPC framework ...see more

Sigurdur Magnus Gardarsson and 26 others
 1 repost

Like Comment Repost Send

922 impressions View analytics

NCC Iceland: TrustLLM use Case with SME
Mideind ehf on LLM on HPC

NEW

EuroHPC JU Hosting Sites

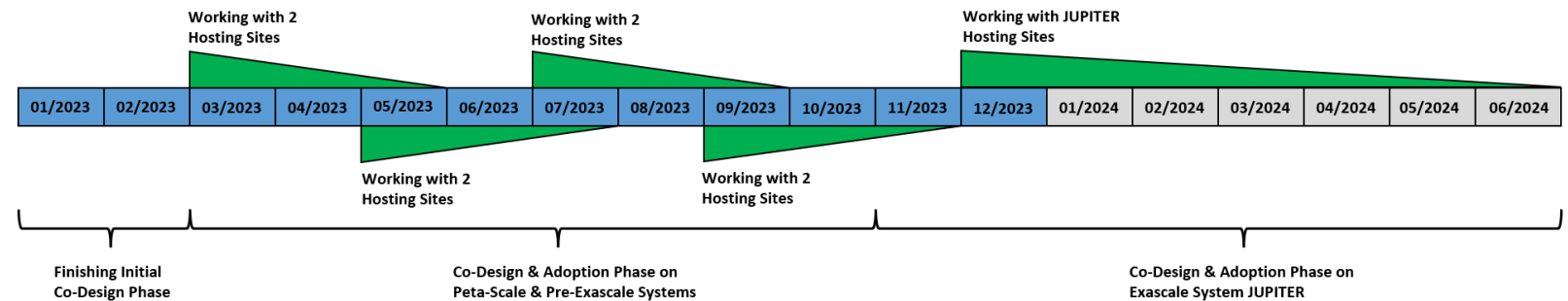
- LUMI
- VEGA
- KAROLINA
- DEUCALION
- LEONARDO
- MELUXINA
- DISCOVERER

EuroHPC

Apptainer Container Environment

BSC-CNS HPC System
 Mare Nostrum

JUPITER Exascale System



NCCs & Industry Hosting Sites – Adoption Plans

NEW



TBD: Computing Time Grants, Karolina, Meluxina

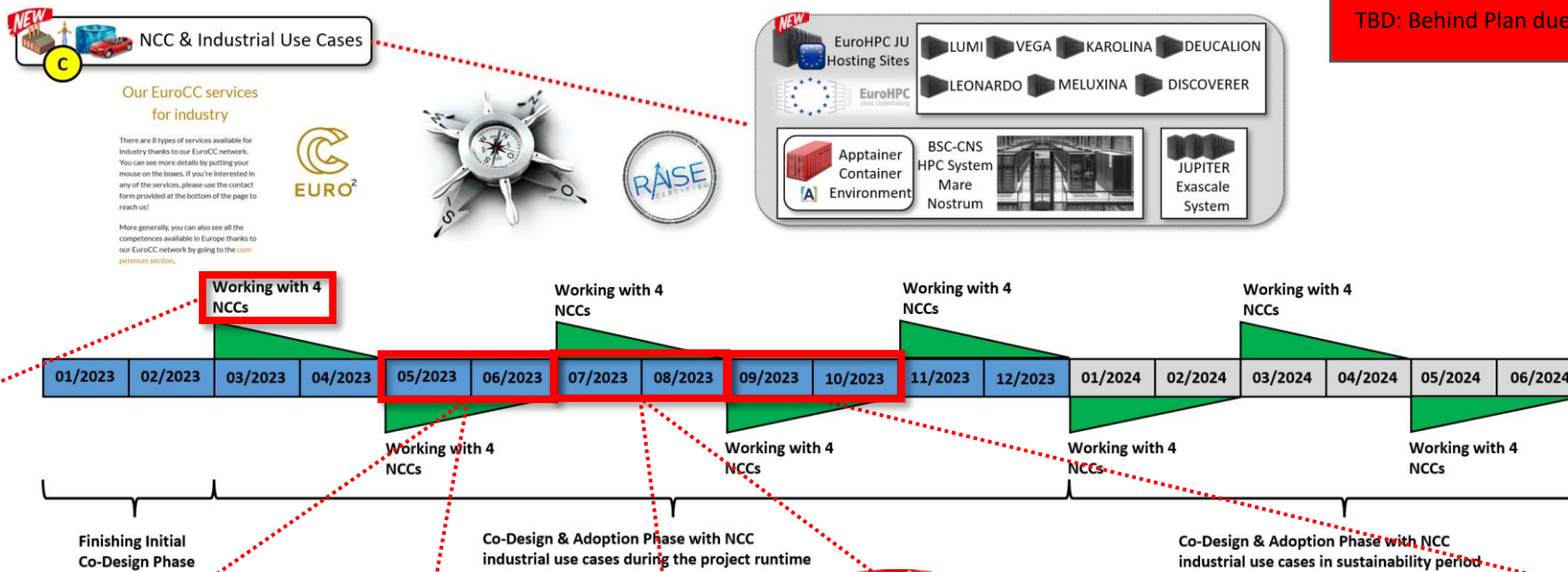
TBD: Behind Plan due to vacation time in June/July

NCC Iceland
(Gothenburg, industrial use case Mideind SME)

NCC Sweden
(Gothenburg, Telco Meeting on 2023-03-28, Documents & Material analysis)

NCC Czech Republic
(Gothenburg)

NCC Luxembourg
(EuroCC Review, Telco Meeting on 2023-03-27)



NCC Germany (2023-06-13)

NCC Slovenia

NCC Croatia

NCC North Makedonia (2023-05-24)

NCC Belgium (2023-06-13)

Vacation Time – no other meetings

July CASTIEL Code of the Month Internal Event for NCCs (2023-07-19)

NEW

August: NCC Iceland meeting with Mideind on LLM activities & use case (2023-08-24)

NCC Spain : BSC Icewind SME AI & CFD; Mideind ehf NLP & OpenAI

NCC Austria: Interested, more info, meeting after Easter

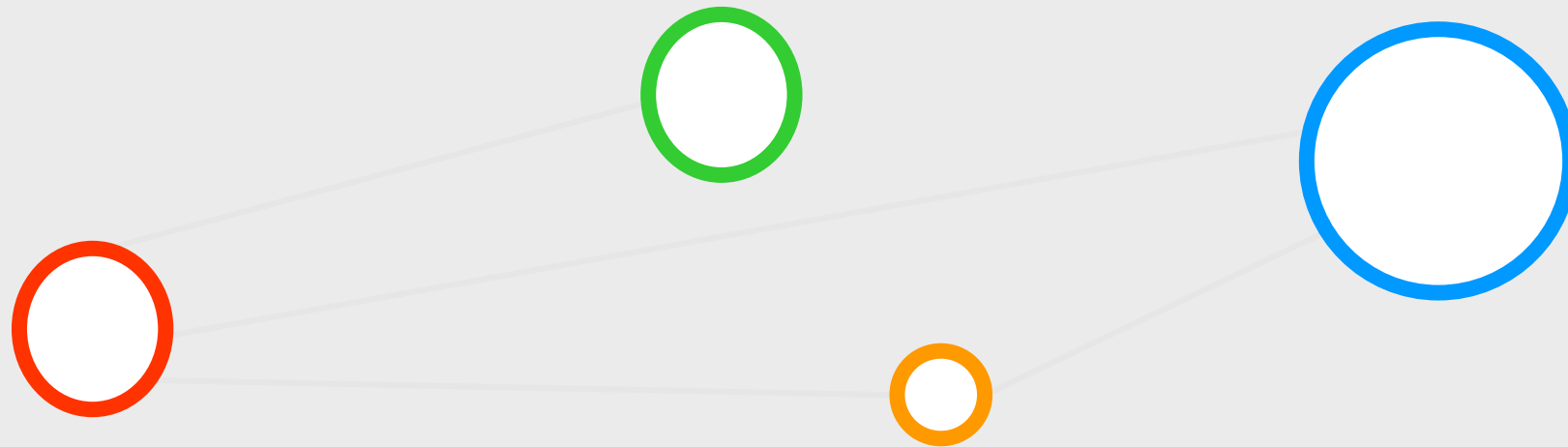
NCC Greek

NCC Latvia → Industry? Industry HPC not much, AI use case, Lauris might now it, biomedical applications

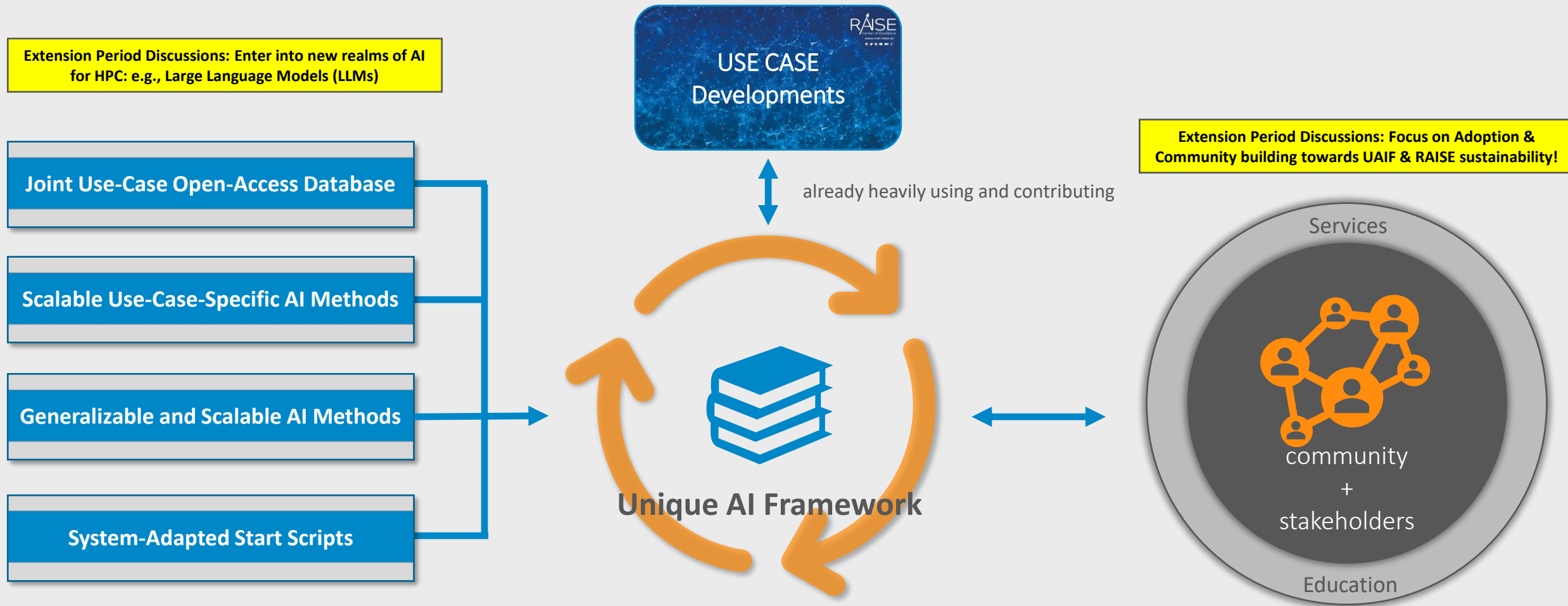
ATOS maybe, active in many



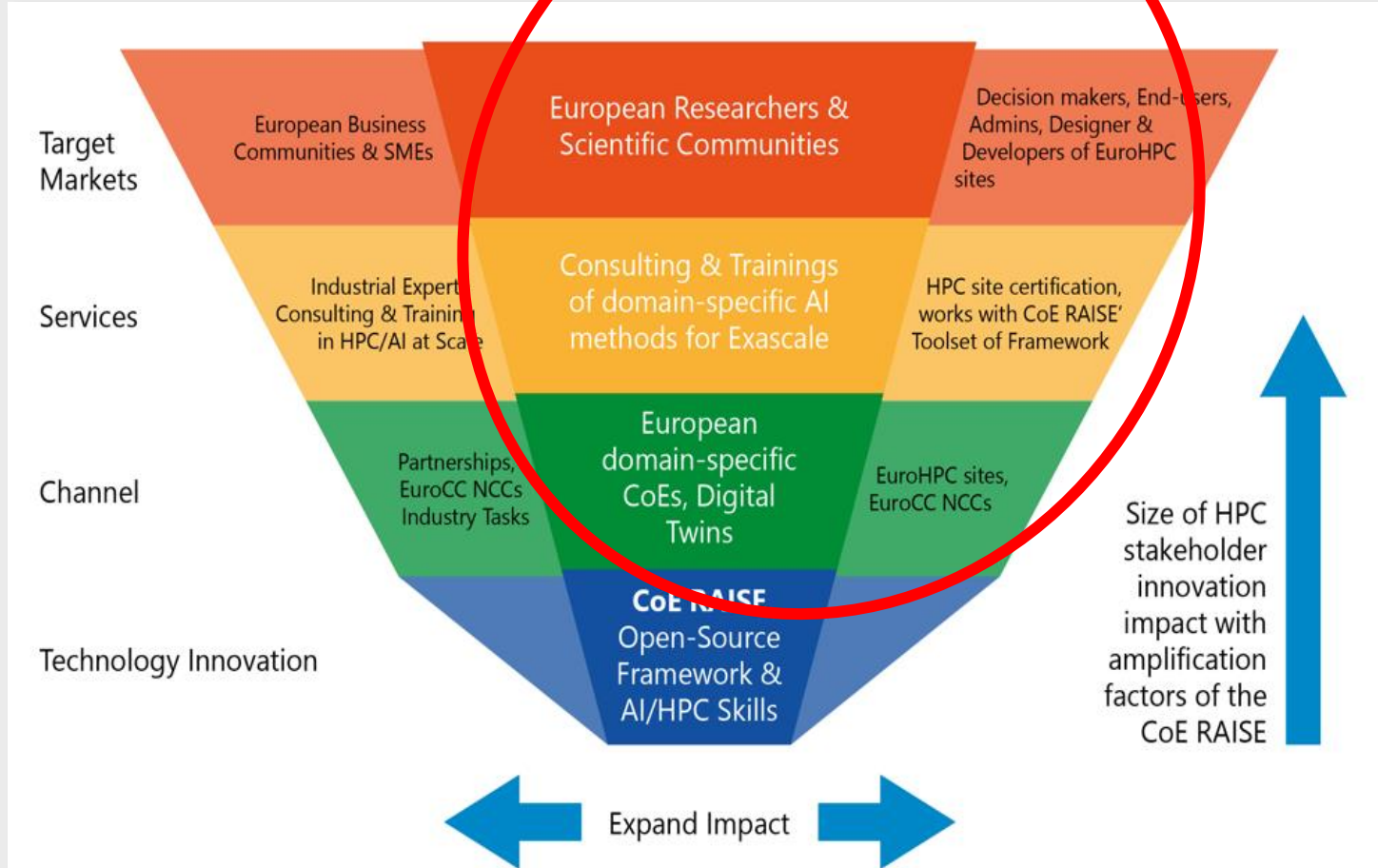
Summary & Q&A



Summary: Unique AI Framework Overview



Q&A: NCC +COEs Adoption Feedback?



LUMI

LUMI is a pre-exascale EuroHPC supercomputer located in Kajana, 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	550 petaflops Peak 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 XH2000 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 petaflops Sustained performance	10,05 petaflops Peak performance
--	--

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

12,81 petaflops Sustained performance	18,29 petaflops Peak performance
---	--

KAROLINA

Karolina is a petascale EuroHPC supercomputer located in Olšava, 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
--	--

drive. enable. innovate.



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

Follow us:



R^G