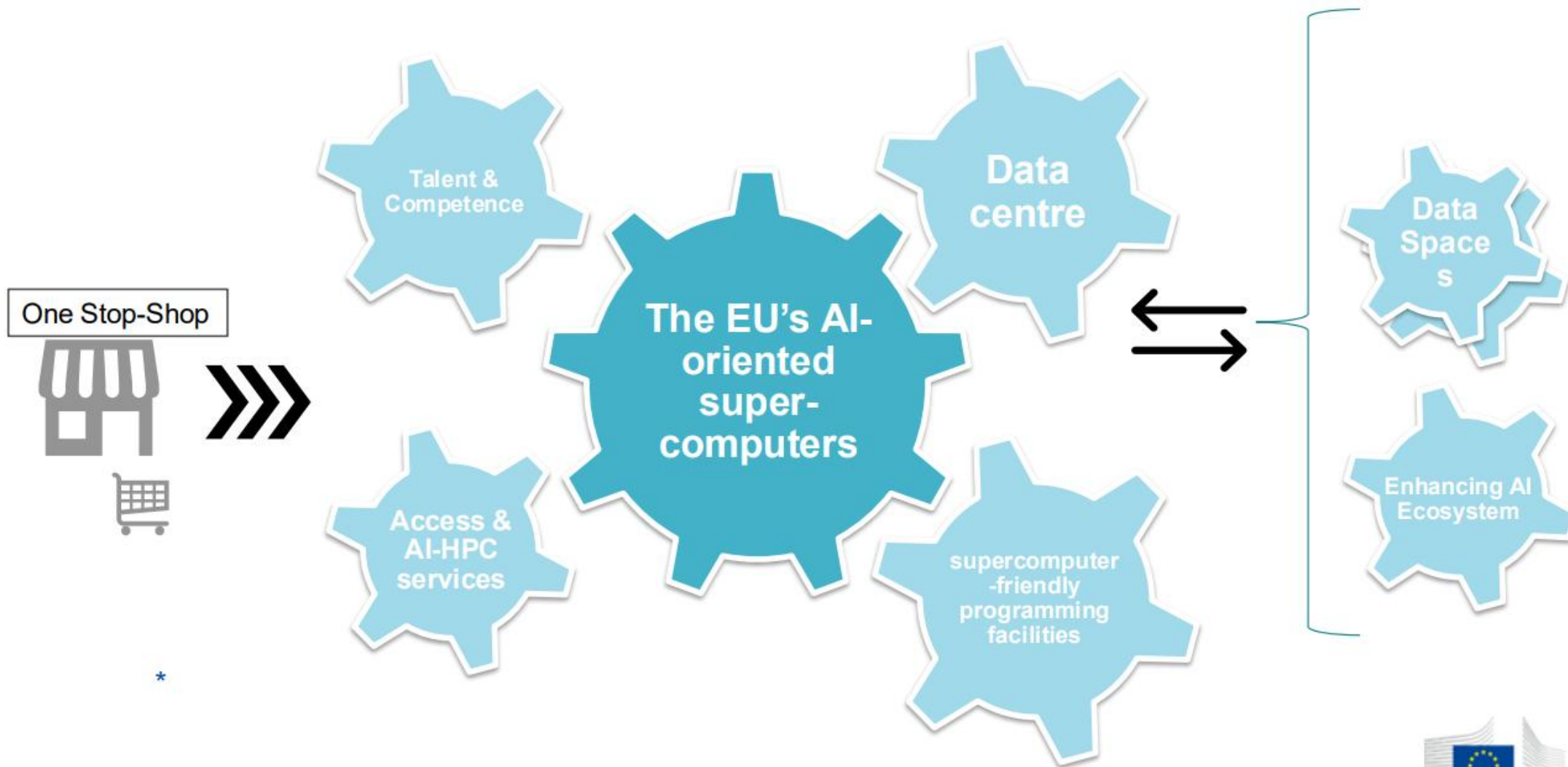# Services in the AI Factories: Empowering Innovation Across Europe

Laura Morselli – CINECA
IT4LIA Technical Project Manager
l.morselli@cineca.it

The pillars of AI Factories

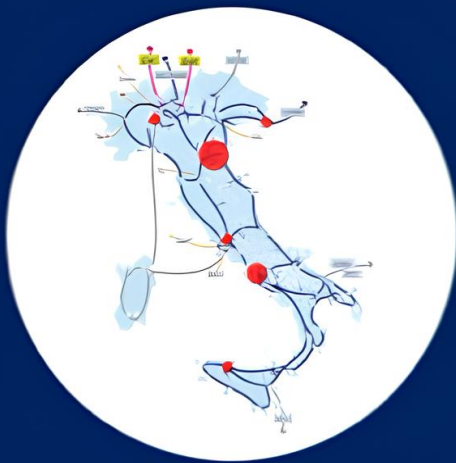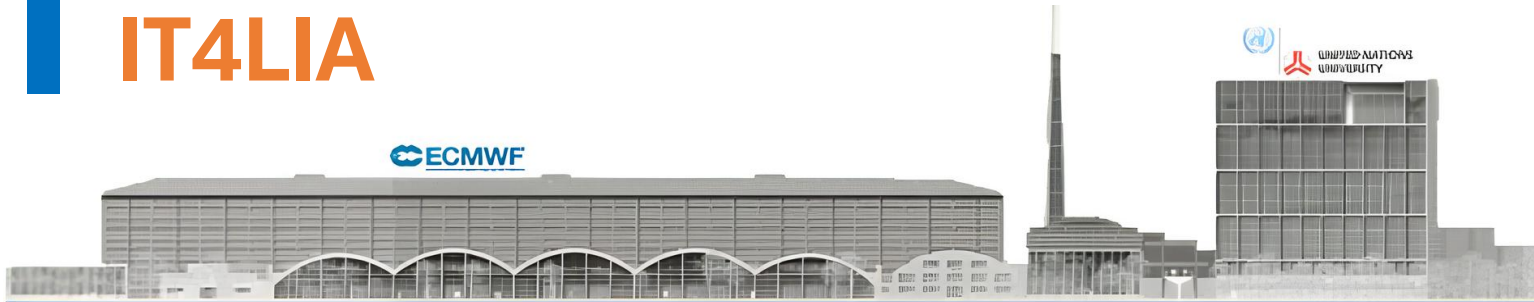# Activities of AI Factories

- **Compute:** demonstrate a minimum <u>available</u> or <u>planned</u> AI-optimized computing capacity required for supporting an AI Factory ecosystem
- **Storage:** co-located with the supercomputer / high-speed / capacity storage
- **Data:**
  - Availability and access to high-quality data repositories
  - Data centre facility co-located with supercomputer
  - Identify access to selected Common European Data Spaces
  - Security: Protection of sensitive data
  - Secure Trusted Environments: results are usable & trustable (e.g. compliant with AI Act)
- **Connectivity:** High-bandwidth, low-latency secure networking
- **Software & Application Development:** Ready-to-use set of AI-oriented tools

Activities of AI Factories

- **User Support:** support activities to be offered to users (e.g., optimising the computing usage to speed up training, optimising the model for efficient deployment); 24/7 service

- **Skills / Education:** Skills Plan outlining the skills needed for the AI stakeholders (e.g. HPC advanced courses (deep learning, AI programming environments, hackathons, traineeships)

- **Engagement / Interacting with AI community:** Interacting with AI-ecosystem at large & other AI initiatives (e.g., TEFs, EDIHs, NCCs)

- **Guaranteeing trustworthy AI:** Collaboration with the AI office

- **AI Factories networking:** A collaborative framework to ensure an effective networking & resource optimisation among AI Factories

ALL ELEMENTS SHOULD BE ALIGNED AND MUTUALLY REINFORCE EACH OTHER

The Italian AIF IT4LIA

The Italian AIF
IT4LIA

AI FACTORY USERS

AI for Science
Academia, Research

AI for Innovation
Startups, SMEs, Industries, Private Entities

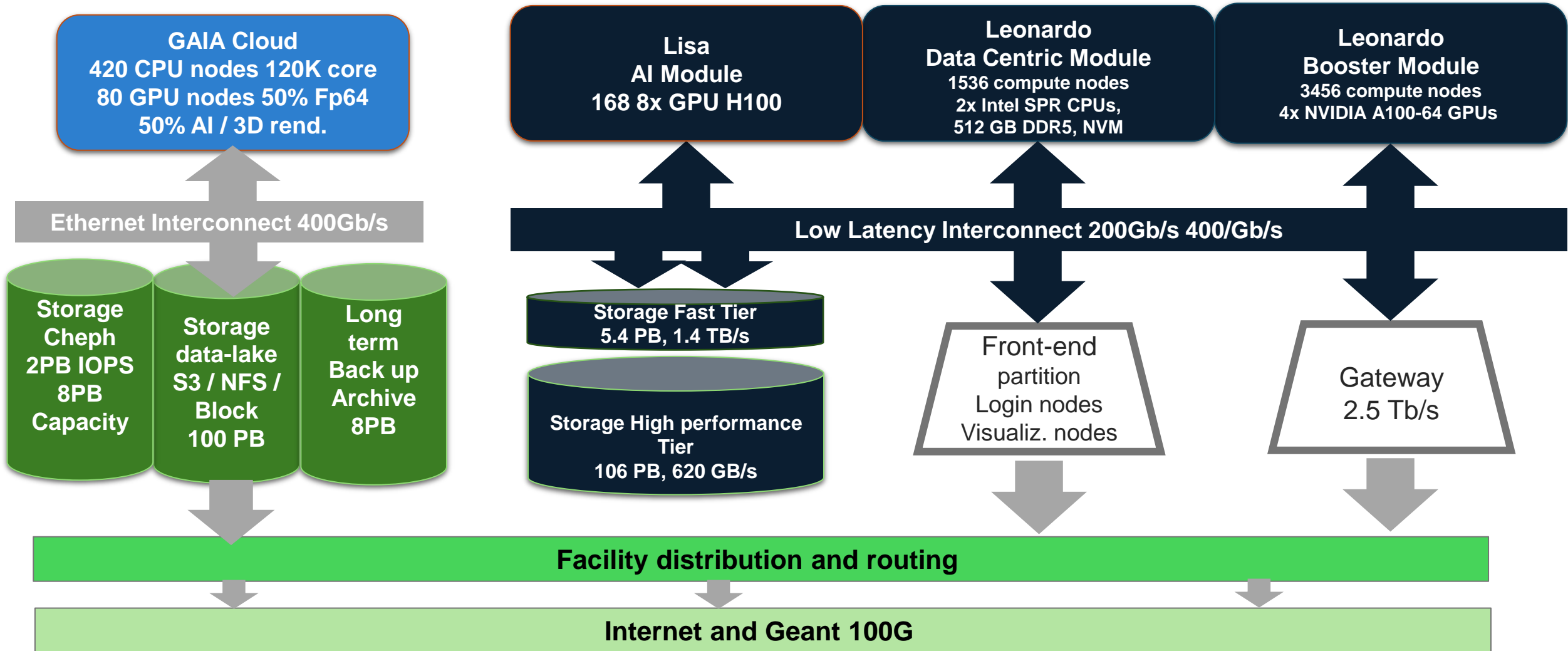### AI for Public Good
Public Administrations

AI FACTORY SERVICES

- 50% for the MS, 50% for EU Users
- Free for Academia, PA, start-ups, SMEs
- Access to compute, access to data, expert consultancy, training, ...

AI OPTIMIZED SUPERCOMPUTER(S)

THE AI OPTIMIZED SUPERCOMPUTERS
LEONARDO + LISA + GAIA CLOUD

GAIA Cloud
420 CPU nodes 120K core
80 GPU nodes 50% Fp64
50% AI / 3D rend.

Lisa
AI Module
168 8x GPU H100

Leonardo
Data Centric Module
1536 compute nodes
2x Intel SPR CPUs,
512 GB DDR5, NVM

Leonardo
Booster Module
3456 compute nodes
4x NVIDIA A100-64 GPUs

Ethernet Interconnect 400Gb/s

Low Latency Interconnect 200Gb/s 400/Gb/s

Storage
Cheph
2PB IOPS
8PB
Capacity

Storage
data-lake
S3 / NFS /
Block
100 PB

Long
term
Back up
Archive
8PB

Storage Fast Tier
5.4 PB, 1.4 TB/s

Storage High performance
Tier
106 PB, 620 GB/s

Front-end
partition
Login nodes
Visualiz. nodes

Gateway
2.5 Tb/s

Facility distribution and routing

Internet and Geant 100G

→ Ethernet

→ InfiniBand

Leonardo
Lisa

GAIA Cloud

CINECA

The Services of IT4LIA

DATA-RELATED SERVICES
ACCESS – TRANSFER – MANAGEMENT – GENERATION

ACCESS TO ON-PREM DATASETS

EXTERNAL DATA ACQUISITION

DATA ORCHESTRATION

CACHE-MEDIATED DATA ACCESS

MULTIPROTOCOL STORAGE SERVICE

METADATA ENGINE SERVICE

SYNTHETIC DATA GENERATION

The Services of IT4LIA

DATA-RELATED SERVICES	HORIZONTAL SERVICES
ACCESS – TRANSFER – MANAGEMENT – GENERATION	AI SETUP – AI DEVELOPMENT – AI TRUST – AI TEST – AI NETWORK
ACCESS TO ON-PREM DATASETS	EXPERT CONSULTANCY FOR AI on HPC: PRE-TRAINING, DOMAIN ADAPTATION, FINE TUNING, RAG, MODEL COMPRESSION AND OPTIMIZATION, AGENTS, …
EXTERNAL DATA ACQUISITION	
DATA ORCHESTRATION	RAPID SETUP OF AI FRAMEWORKS AND WORKFLOWS
CACHE-MEDIATED DATA ACCESS	ACCESS TO ON PREM MODELS HUB
MULTIPROTOCOL STORAGE SERVICE	FEDERATED LEARNING as a SERVICE
METADATA ENGINE SERVICE	OPEN CALLS FOR SMEs/START-UPS FOR PROJECT CO-DEVELOPMENT
SYNTHETIC DATA GENERATION	STRESS and A/B TESTING OF MODELS
	EVALUATION OF MODEL SAFETY
	SUPPORT ON COMPLIANCE WITH REGULATIONS AND ACCESS TO REGULATROY SANBOXES

The Services of IT4LIA

DATA-RELATED SERVICES
ACCESS – TRANSFER – MANAGEMENT – GENERATION

- ACCESS TO ON-PREM DATASETS
- EXTERNAL DATA ACQUISITION
- DATA ORCHESTRATION
- CACHE-MEDIATED DATA ACCESS
- MULTIPROTOCOL STORAGE SERVICE
- METADATA ENGINE SERVICE
- SYNTHETIC DATA GENERATION

HORIZONTAL SERVICES
AI SETUP – AI DEVELOPMENT – AI TRUST – AI TEST – AI NETWORK

- EXPERT CONSULTANCY FOR AI on HPC: PRE-TRAINING, DOMAIN ADAPTATION, FINE TUNING, RAG, MODEL COMPRESSION AND OPTIMIZATION, AGENTS, …
- RAPID SETUP OF AI FRAMEWORKS AND WORKFLOWS
- ACCESS TO ON PREM MODELS HUB
- FEDERATED LEARNING as a SERVICE
- OPEN CALLS FOR SMEs/START-UPS FOR PROJECT CO-DEVELOPMENT
- STRESS and A/B TESTING OF MODELS
- EVALUATION OF MODEL SAFETY
- SUPPORT ON COMPLIANCE WITH REGULATIONS AND ACCESS TO REGULATROY SANBOXES

VERTICAL SERVICES
AGRIFOOD – CYBERSECURITY – EARTH – MANUFACTURING

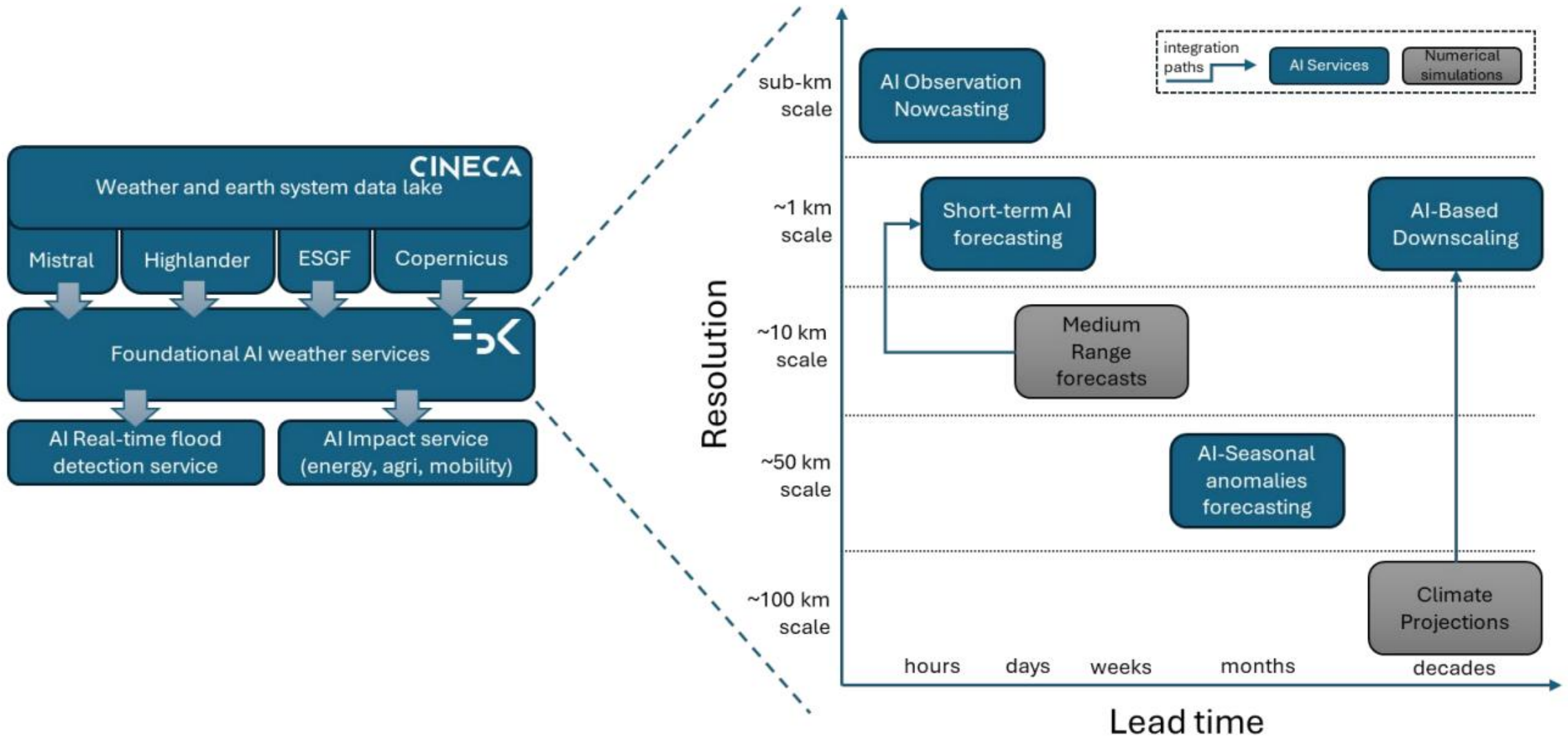- DEVELOPMENT OF OPEN-SOURCE, COMPLIANT, DOMAIN-SPECIFIC LARGE-SCALE MODELS
- CURATION OF HIGH-QUALITY, COMPLIANT, SECTORIAL DATASETS
- FINE-TUNING AND DEVELOPMENT OF SERVICES FOR THE ECOSYSTEM BASED ON THE LARGE-SCALE MODELS
- DEPLOYMENT OF TAILORED AI SERVICES
- DOMAIN-SPECIFIC EXPERT CONSULTANCY

The Services of IT4LIA

The Services of IT4LIA

DATA-RELATED SERVICES
ACCESS – TRANSFER – MANAGEMENT – GENERATION

- ACCESS TO ON-PREM DATASETS
- EXTERNAL DATA ACQUISITION
- DATA ORCHESTRATION
- CACHE-MEDIATED DATA ACCESS
- MULTIPROTOCOL STORAGE SERVICE
- METADATA ENGINE SERVICE
- SYNTHETIC DATA GENERATION

HORIZONTAL SERVICES
AI SETUP – AI DEVELOPMENT – AI TRUST – AI TEST – AI NETWORK

- EXPERT CONSULTANCY FOR AI on HPC: PRE-TRAINING, DOMAIN ADAPTATION, FINE TUNING, RAG, MODEL COMPRESSION AND OPTIMIZATION, AGENTS, …
- RAPID SETUP OF AI FRAMEWORKS AND WORKFLOWS
- ACCESS TO ON PREM MODELS HUB
- FEDERATED LEARNING as a SERVICE
- OPEN CALLS FOR SMEs/START-UPS FOR PROJECT CO-DEVELOPMENT
- STRESS and A/B TESTING OF MODELS
- EVALUATION OF MODEL SAFETY
- SUPPORT ON COMPLIANCE WITH REGULATIONS AND ACCESS TO REGULATROY SANBOXES

VERTICAL SERVICES
AGRIFOOD – CYBERSECURITY – EARTH – MANUFACTURING

- DEVELOPMENT OF OPEN-SOURCE, COMPLIANT, DOMAIN-SPECIFIC LARGE-SCALE MODELS
- CURATION OF HIGH-QUALITY, COMPLIANT, SECTORIAL DATASETS
- FINE-TUNING AND DEVELOPMENT OF SERVICES FOR THE ECOSYSTEM BASED ON THE LARGE-SCALE MODELS
- DEPLOYMENT OF TAILORED AI SERVICES
- DOMAIN-SPECIFIC EXPERT CONSULTANCY

AI SKILLS
TRAINING – CAPACITY BUILDING – UP-SKILLIN & RE-SKILLING

- AI FACTORY LEARNING PLATFORM
- LIVE REMOTE SESSIONS
- SELF-PACED COURSES
- HACKATHONS
- FACE-TO-FACE TRAINIGS
- ADVANCED SCHOOLS
- INTERNSHIPS
- LIVE Q&A SESSIONS WITH EXPERTS
- HANDS-ON ON REAL PROBLEMS
- CERTIFICATIONS AND CAREER DEVELOPMENT

Few "take home" messages

- Starting from April/May, new opportunities will be available for **SMEs and start-ups** to access **AI-optimized HPC resources**—both at the national and EU levels. These add to the existing access already available for academia and public administrations.

- Developing AI models and AI-powered services becomes easier thanks to a **comprehensive ecosystem of tools, infrastructure, and expert support**.

- Some services, particularly those involving expert consultancy, will be available **immediately**. Others will be developed internally and progressively released to the ecosystem.

IT4LIA TAKES OFF TOMORROW!

Laura Morselli – CINECA
IT4LIA Technical Project Manager
l.morselli@cineca.it
superc@cineca.it