Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

ICSC
Centro Nazionale di Ricerca in HPC,
Big Data and Quantum Computing

~~Why the SOSC, after all?~~ What to do after the SOSC?

Davide Salomoni, ICSC Innovation Manager – davide@supercomputing-icsc.it

SOSC 2024, Bologna 2-6 December 2024

A bit of history of these presentations

2017

Introducing Cloud Computing and OpenStack

Davide Salomoni, INFN-CNAF
School on Open Science Cloud
Perugia, 5-9/6/2017

THE ANSWER IS 42

2018

Opportunities *for you* (2)

- If you are interested in research associate / post-doc positions, master / bachelor thesis or short-term contracts at INFN (multiple sites) or at other institutions who contributed to this School, please send email to sosc18@fisgeo.unipg.it and we'll get you in touch with the appropriate people.
- Info from V. Kuznetsov: see https://sites.google.com/view/valentin-kuznetsov/projects for a list of his R&D projects.
- For what regards CERN-related programs, see
  - https://home.cern/students-educators/summer-student-programme
  - https://home.cern/students-educators/updates/2017/12/applications-open-2018-cern-openlab-summer-students
  - http://care
  - https://car

Davide Salomoni

2019

Opportunities *for you* (2)

- A list of Valentin's R&D projects is available at https://sites.google.com/view/valentin-kuznetsov/projects. Get in touch with him if you want to know more about them.
- Various opportunities available at INFN CNAF (Bologna) are listed at https://www.cnaf.infn.it/tirocini-tesi-assegni/. Get in touch with Davide or with the people listed in that page if you want to know more about them.
  - Note in particular the following two-year contract positions for "assegni di ricerca" currently open at CNAF:
    - http://www.ac.infn.it/job/dettagli_job.php?id=2578
    - http://www.ac.infn.it/job/dettagli_job.php?id=2591
    - http://www.ac.infn.it/job/dettagli_job.php?id=2592

SOSC 2019 Wrap-up and Opportunities 4

2022

Opportunity areas

- Collaborations with INFN related to computing topics mentioned at SOSC22 are possible in technical areas such as:
  - System administration
  - (Advanced) cloud computing infrastructures, including integration with HPC, GPU, FPGA, adaptation of multi-disciplinary use cases
  - User support
  - Middleware development
  - Security / data protection
  - Scientific communication
- The detail of the technologies involved in these opportunities is too long to list. Have a look at topics covered at SOSC22 for a sample.

Davide Salomoni SOSC22 - Perugia, 2/12/2022 2

So, history tells us that…

- "**Opportunities**" is the main topic of these presentations.
  - What *type* of opportunities? → to obtain know-how, resources, services, or a new job position ☺
- Notably, starting from 2023, the "Opportunities" are not confined anymore to INFN only, but span the larger ICSC ecosystem.

To present the opportunities in a meaningful way, it is important to recap where these opportunities come from and what the overall vision is.

On Monday we said…

Toward the Future…

(~~not now~~ – we'll discuss this ~~on Friday~~ today)

2018

33
zettabytes

Salvati su tablet da 512 GB, formerebbero una torre che arriva fino alla Luna

x 2 nel 2021

5.3 x

2025

175
zettabytes

Sufficienti per andare sulla Luna e tornare cinque volte

2018

301
miliardi

2.4 % of EU GDP

2.8 x

2025

829
miliardi

5,8 % of EU GDP

2018

5,7
milioni

professionisti dei dati

1.9 x

2025

10,9
milioni

professionisti dei dati

…

The third pillar is guiding innovation in a responsible way.

Thanks to our investment in the last years, Europe has now become a leader in supercomputing – with 3 of the 5 most powerful supercomputers in the world.

We need to capitalise on this. This is why I can announce today a new initiative to open up our high-performance computers to AI start-ups to train their models.

…

EU State of the Union – Strasbourg, 13 September 2023

2. Industry takes the lead

Until 2014, academia led in the release of machine learning models. That's no longer the case. In 2023, there were 51 machine learning models produced by industry compared with just 15 from academia. Interestingly, 21 notable models were created in 2023 as a result of industry-academia collaborations, which represents a new high.

What's behind industry's phenomenal uplift? Creating cutting-edge AI models now demands a substantial amount of data, computing power and financial resources, which are not typically accessible in academia.

3. Frontier models reach unprecedented costs

As mentioned before, LLMs aren't cheap to run or train. According to AI Index estimates, the training costs of leading AI models have increased significantly. For example, OpenAI's GPT-4 training costs were estimated to be $78 million, while Gemini Ultra by Google cost $191 million.

In comparison, back in 2017, the original Transformer model, which is recognized as introducing the architecture that underpins virtually all modern LLMs, cost around $900 to train.

From the **Stanford University AI Index Report**, 2024 – https://aiindex.stanford.edu/report/ (502 pages!)

https://www.aboutamazon.com/news/aws/ai-cloud-adoption-economic-impact-gdp-aws

https://tinyurl.com/25ebd4db
(Part A)

https://tinyurl.com/a8v3h26m
(Part B)

From the report "The future of European competitiveness" – 1

- [Increase] computational capacity and [make] available its network of high-performance computers. While so far this capacity has been mostly used for scientific research, the Commission is progressively opening it to AI start-ups, SMEs and the broader AI community.

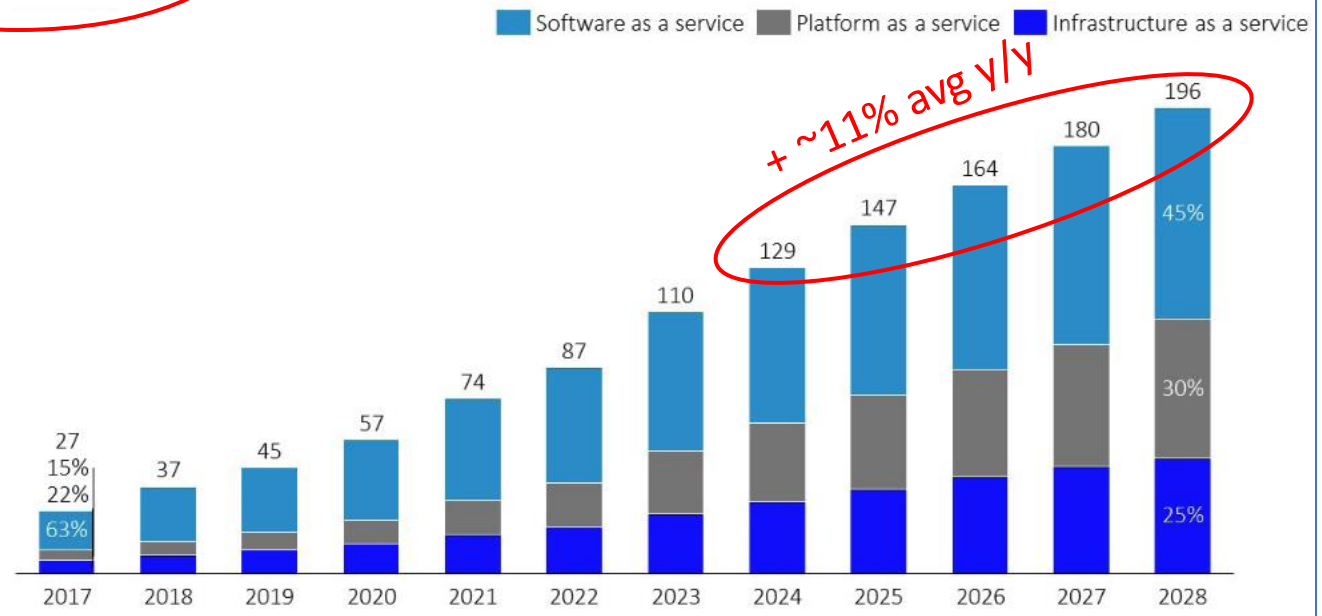
Facts

- On the "EU cloud services market": most EU providers offer basic services in the form of infrastructure-as-a-service (IaaS) and mostly depend on hosting or re-selling hyperscalers' platform services (PaaS), which are harder to compete with, commercially stickier and more profitable.

FIGURE 6

EU cloud market size

EUR billion



Source: Statista Technology Market Insights, 2024.

The global HPC market was valued at USD 48.5 billion in 2022 and is estimated to grow at a compound annual growth rate (CAGR) of 7.5% between 2023 and 2030

From the report "The future of European competitiveness" – 2

- The EU has secured a strong international position in high-performance computing (HPC) – a unique advantage to exploit in areas such as AI, and to stimulate private investment.
- AI developments are an opportunity for EU industrial players to boost their competitiveness **but also a risk** to lose their leadership and profitability if AI is not rapidly integrated in their offerings.
- Quantum computing, the next trailblazing innovation in the computing field, could open new opportunities for the EU's industrial competitiveness and technological sovereignty.

Facts

Some of the recommendations of the report

- Regularly **increase computational capacity** dedicated to the *training* and *algorithmic development of AI models* in existing EU HPC centers.
- Finance the **expansion of Euro-HPC to additional cloud and storage capabilities to support AI training and extend their activity to AI fine-tuning and inference**.
- Open up Euro-HPC to **a 'federated AI model' favoring cooperation of public-private infrastructure** to provide AI training power, leveraging the joint capacity of public computing and private resources and increasing the EU's competitive scale.
- Develop **quantum labs or nodes attached to all EU HPC centers and launch public-private partnerships** – involving large EU tech leaders as a priority – to co-invest in the whole frontier tech stack, including neuro-morphic and quantum chips.
- Leverage the EU-wide **coordination and harmonization of national AI sandbox regimes**, and ensure harmonized and simplified implementation of the GDPR

PRESS RELEASE | 4 December 2024 | European High-Performance Computing Joint Undertaking | 2 min read

The Innovate Consortium Chosen to Host EuroHPC's First Industrial Supercomputer

The Innovate consortium, led by CINECA and including seven Italian industrial partners from diverse sectors, has been selected to host and operate the first EuroHPC industrial-grade supercomputer in Bologna, Italy.

GoodandEvil - stock.adobe.com

This EuroHPC industrial-grade supercomputer will provide supercomputing capabilities, specifically tailored to meet the security, confidentiality and data integrity needs of European industrial users. Managed and operated by the Innovate consortium, this versatile HPC system will support a wide range of industrial applications. As the first EuroHPC system of its kind, this new system will play a key role in enhancing the innovation potential of European enterprises, strengthening their know-how and fostering the developing advanced computing skills.

This EuroHPC industrial-grade supercomputer will provide supercomputing capabilities, specifically tailored to meet the security, confidentiality and data integrity needs of European industrial users.

https://eurohpc-ju.europa.eu/innovate-consortium-chosen-host-eurohpcs-first-industrial-supercomputer-2024-12-04_en

EUSAir: EU Regulatory Sandboxes for AI

EUSAir is an EU project for building AI sandboxes and related guidelines.

AI regulatory sandboxes are closed test environments where innovations can be developed, tested and validated. They will enable developers and providers of AI applications and solutions to verify their regulatory compliance before they are placed on the market and avoid fines for breaching the EU Artificial Intelligence Act.

The EUSAir project will develop EU-wide guidance on the implementation of AI regulatory sandboxes in line with the EU Artificial Intelligence Act.

ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

NEWS ARTICLE | 1 August 2024 | Directorate-General for Communication | 2 min read

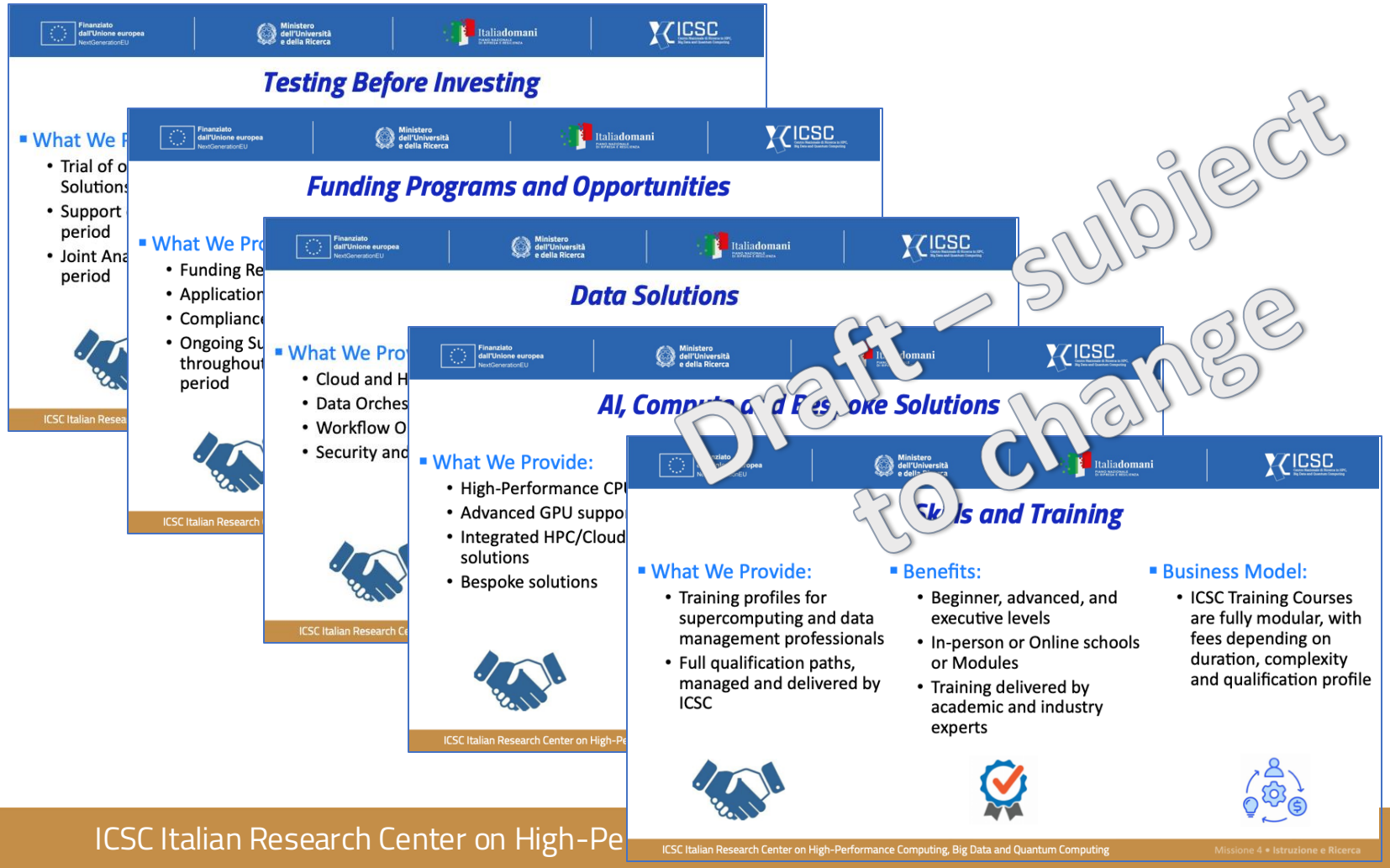
AI Act enters into force

On 1 August 2024, the **European Artificial Intelligence Act (AI Act) enters into force**. The Act aims to **foster responsible artificial intelligence development and deployment in the EU**.

ICSC – The HPC, Big Data, Quantum *one-stop-shop*

Preview of the upcoming ICSC solution portfolio

For enterprises and public institutions



Draft – subject to change

The ICSC Solution Portfolio 2025

- Testing Before Investing
- Funding Programs and Opportunities
- Data Solutions
- AI, Compute and Bespoke Solutions
- Skills and Training

Some of the recommendations of the report

- Regularly increase computational capacity dedicated to the *training* and *algorithmic development* of massive AI models
- Finance the expansion of Euro-HPC to additional cloud and storage capabilities **to support AI** training and experimentation and inference.
- Open up Euro-HPC to a **'federated AI model' favoring cooperation of public-private** infrastructure to provide AI training power leveraging the joint capacity of public computing and private clouds and increasing the EU's competitive scale.
- Develop quantum computing in the EU and launch **public-private partnerships** – involving large EU tech leaders as a priority – to co-invest in the whole frontier tech stack, including neuro-morphic and quantum chips.
- Leverage the EU-wide coordination and harmonization of national AI **sandbox regimes**, and ensure harmonized and simplified implementation of the GDPR

It is up to us all to make this a full reality. If interested in these topics and in this vision, come on board!(*)

(*) Open **job positions are available** at the ICSC Foundation and at INFN.
If interested, contact sosc24-pc@lists.infn.it


ICSC
Centro Nazionale di Ricerca in HPC,
Big Data and Quantum Computing
Supercomputing shaping the future
https://www.supercomputing-icsc.it/
e-mail: info@supercomputing-icsc.it