



New milestone for the French AI flagship, Jean Zay, to drive major AI breakthroughs, thanks to a record Eviden-led installation time of its new NVIDIA Hopper partition

Atlanta, 11/20/2024

The leading French supercomputer Jean Zay, addressing the convergence between HPC and AI, has been extended for the 4th time since 2019 to sustain very high demand from the French AI community and foster the rise of generative AI across many disciplines and companies, especially French AI startups. This is fully aligned with the vision of the French President's Emmanuel Macron, during VIVATECH 2023, to strengthen the French #AIForHumanity strategy.

In order to answer to the strong demand from the French AI community (>1000 research projects in AI supported in 2023) and the rise of Generative AI in France, GENCI has received 40M€ funding to increase the capacities of its French AI flagship Jean Zay, a supercomputer hosted and operated by Institute for Development and Resources in Intensive Scientific Computing (IDRIS - CNRS) providing tailored AI user-support. With that, the new computing partition, from the French provider Eviden, holds a total of 1,456 NVIDIA Hopper GPUs hosted in 14 [BullSequana AI 1200H](#) racks, 364 direct liquid cooling blades, each one having 2 Intel CPUs, 4 NVIDIA Hopper SXM 80GB GPUs and 4 [NVIDIA ConnectX-7](#) 400 Gbps InfiniBand adapters, connecting to NVIDIA [Quantum-2](#) InfiniBand switches. In addition, a renewed tiered storage of 4.3 PB flash drives from DDN has been delivered providing more than 1.2 TB/s of Read/Write bandwidth to sustain I/O intensive AI workloads and close to 40 PB of high-speed rotating disks, all together using a Lustre filesystem.

The extension was awarded to Eviden in March 2024 with a 4-month record-time installation allowing the new partition to already serve 13 Grand Challenges from July 2024. During a 3 months warm-up phase, these 13 scientific projects had dedicated access to the full extension capacity with tight joint expertise support from IDRIS, Eviden and NVIDIA to demonstrate scientific and industrial breakthroughs in the areas of AI, AI4S (AI For Science) and quantum emulation using hundreds of GPUs.

“The Jean Zay supercomputer is a crucial milestone to boost French AI research and to bring together France’s academic and industrial research communities,” commented Bruno Lecointe, VP, Global head of HPC, AI and Quantum Computing at Eviden, Atos Group. *“Eviden is immensely proud to support the GENCI and CNRS in tackling their AI challenges and to have been able to deliver a key element in France’s technological competitiveness in such a short timeframe. We look forward to seeing the breakthroughs enables by Jean Zay and to deepening our collaboration together.”*

Collaboration on the 13 Grand Challenges includes:

- 8 industrials challenges: Projects are from leading AI startups among which Hugging Face, H-Company, Owkin, Bioprimus, Linagora, Pleias and Zaion and one is performed by Valeo, a large industrial company through its Valeo.AI open lab.
- One additional industrial project from Qubit Pharmaceuticals is related to the unprecedented scaling of a quantum emulator used for simulating quantum chemistry workloads in materials sciences on up to 290 qubits.
- 4 academic challenges: CNRS with Flatiron Institute, Inria, Sorbonne University and University of Lorraine.

Among the 13 Grand Challenges, 6 are related to AI4S in materials science, chemistry, astrophysics or life sciences while the others are focusing on the development of sovereign open source LLMs, from large action models to improving the vulnerability of recent NLP models to attacks, through models serving public administration and large companies concerned with compliance & ethics with regard to environmental impact.

While these Grand challenges are still progressing, it is already possible to highlight some few preliminary outcomes achieved thanks to the new Jean Zay NVIDIA Hopper partition:

- **LINAGORA**, partnering with the OpenLLM-France consortium, has trained LUCIE a new 7B and open-source sovereign multi-modal foundation model using 512 GPUs over 3 months, on Jean Zay supercomputer, with the main goal to support the generation of educational content, first in French and later for 4 additional European languages as coding/mathematics problems using a 3T token high quality database and a new tokenizer with SOTA fertility. This open-source model, will be a cornerstone for creating diverse, high-quality content that supports the next generation of learners in Europe. *“The training of LUCIE marks a key step towards sovereign AI, specifically designed to enhance educational experiences. With LUCIE, LINAGORA and the OpenLLM-France consortium are committed to delivering a truly open-source foundation model, which is a significant milestone in AI for education that can scale across multiple languages and domains, including coding and mathematics. Alongside developing other Small Language Models (SLM), LUCIE figures in the next evolution for Personal Language Models, positioning LINAGORA and the OpenLLM-France consortium at the forefront of the AI revolution.”* said Michel-Marie Maudet, Linagora General Manager
- **Owkin** is the first end-to-end AI techbio. By understanding complex biology through cutting-edge AI, Owkin aims at identifying new treatments, de-risk or accelerate clinical trials, and build diagnostic tools to reduce time to impact for patients. As part of the Grand Challenges on the Jean-Zay supercomputer, Owkin is training models at scale with the goal of using these models to improve the robustness of digital pathology. The models trained on more than 64 NVIDIA Hopper GPUs rely either on Generative Adversarial Networks (GANs) or Diffusion and aim at disentangling biological content and texture/staining from digital pathology images. Such models could eventually be used to improve the robustness and performance of Owkin's diagnostic tools. *“The free grant*

allocation computing resources from GENCI and the strong IDRIS and NVIDIA support have been key to allow Owkin delivering in the future better drugs and diagnostics at scale as to train more robust foundation models for digital pathology” said Jean-Baptiste Schiratti, Lead Research Scientist

- Francois Lanusse (**CNRS**), through **Polymathic AI**, a joint consortium supported by the Simons Foundation and Schmidt Sciences Foundation, is developing the first large-scale, open-source multi-modal (multi-band galaxy images from both ground-based and space-based observatories, optical spectra, and time series), transformer-based Foundation Model for astrophysics. While the main purpose of the current work is to bring out the intrinsic physical properties of astrophysical objects, a longer-term objective of this line of work is to build and release models that can leverage shared concepts across disciplines to be used by researchers worldwide. At this point in time, the 20 persons international team has been developing their own specialized tokenizers for these scientific data modalities and the newly built model has been trained on Jean Zay with several checkpoints on almost 500B tokens (50B tokens– 10 epochs) based on a masked generative modeling training approach. The model is still under intensive training phase, with a growing ingestion of data to enlarge its knowledge and accuracy. *“Participating in the Jean Zay Grand Challenge has been an incredible adventure, giving us an opportunity to put into practice an end-to-end approach for the development of scientific foundation models. This starts with data collection and engagement with domain scientists and goes all the way to solving methodological and engineering questions emerging from the development of models at this scale. We are very much looking forward to sharing everything we’ve learned with the community, hopefully making it much easier for others to help accelerate the development and adoption of these methods”* said François Lanusse, CNRS researcher at Astrophysique, Instrumentation, Modélisation (AIM AIM – CEA/CNRS/Université Paris Cité) and guest researcher at the Flatiron Institute.
- The French startup **Pleias** with the support of Etalab (Dinum) is currently training MarLAnne, a 3B multi-lingual SLM, on fully open and copyright-free dataset, to release the 1st model fully compliant with the EU AI act. The model will first serve French administration and later European public services as large companies concerned with compliance & ethics as environmental impact. The training phase is going by chunks on 48+ NVIDIA Hopper GPUs. *“With our 3B SLM, we’ve optimized both training and future inference pipelines thanks to the new Jean Zay partition with NVIDIA Hopper GPUs & software and the support of the IDRIS team - to deliver a powerful yet resource-efficient model - proving that European public services can deploy fully compliant AI solutions without compromising on performance or environmental responsibility”* said Pierre-Carl Langlais, CTO of Pleias.
- **CNRS** is collaborating with 9 major healthcare players in France to accelerate the development of personalized, predictive, preventive and participatory medicine for drug-resistant epileptic and developmental encephalopathies (EDEs). Coordinated by the Imagine Institute and funded to the tune of 9.9 million euros

over 5 years, the INNOV4-ePiK program is one of the 19 winners of the sixth call for projects "Hospital-University Health Research, RHUs" of the France Program 2030. This project aims to develop innovative diagnostic and therapeutic approaches for patients suffering from EDEs. *“Our lab uses proprietary cutting-edge technology that merges AI and computational biophysics to explore and reveal all possible conformations of potassium channels variants at the atomistic scale. We are grateful to GENCI for the grant of computing allocation hours on the new Jean Zay partition which has been key in demonstrating an absolute scaling on Jean Zay up to at least 256 NVIDIA Hopper GPUs. This is pivotal in providing molecular-level insights to design drugs that regulate these channelopathies. We also strongly thank IDRIS and NVIDIA for their support and expertise”* said Mounir Tarek, CNRS Research Director and Daniel Wiczew PhD Student in deep learning molecular dynamics artificial intelligence at University of Lorraine.

- **Qubit Pharmaceuticals and Sorbonne Université** are pleased to announce the publication of "Shortcut to Chemically Accurate Quantum Computing via Density-based Basis-set Correction" in *Communications Chemistry (Open Access: DOI: 10.1038/s42004-024-01348-3)*. Using GENCI's high-performance computing resources, Diata Traoré and colleagues embedded a quantum computing algorithm for chemistry within classical density-functional theory, achieving chemically accurate results while minimizing quantum resource needs. The quantum emulation breakthrough enabled to reach quantitative quantum-chemistry results on molecules that would otherwise require brute-force quantum calculations using hundreds of logical qubits. It holds significant promise for applications in drug design and materials science.

Jean-Philip Piquemal, Chief Scientific Officer at Qubit Pharmaceuticals and Director of the Laboratoire de Chimie Théorique de Sorbonne Université, remarked: *"This paper demonstrates the transformative potential of bridging classical and quantum computing through innovative density-based corrections within a hybrid HPC-QC approach. By minimizing the quantum resources needed for chemically accurate computations, we are making strides toward practical quantum chemistry applications. Our method achieves unprecedented precision with minimal resources, combining the power of high-performance computing and quantum algorithms to tackle complex challenges in drug discovery and materials science."*

While these projects are all undergoing, the new Jean Zay partition based on NVIDIA Hopper GPUs is now in production for a broader audience of AI teams/projects since early October 2024. Jean Zay computing resources are accessible free of charge for open research projects from academia and industry and one of its major role will be to serve the development and the training / finetuning of sovereign multimodal foundation GenAI models in France and in Europe. The objective will be to double the number of AI yearly projects supported by year as well as supporting some strategic projects.

“The Jean Zay supercomputer with NVIDIA accelerated computing will play an important role in advancing AI and HPC research in France and Europe,” said John Josephakis, Global VP of Sales and Business Development for HPC and Supercomputing at NVIDIA. *“Using the NVIDIA Hopper platform for these Grand Challenges showcases the promise for researchers to tackle real-world AI problems that can lead to breakthroughs in science and technology.”*

In 2023 and with the support of IDRIS teams, Jean Zay provided access to more than 1,000 AI open research projects from academia and industry. The high demand for the new partition from industrial players (10% of the projects but close to 33% of the cycles used) leaves no doubt that such systems are crucial to sustain the development of sovereign models for scientific and societal challenges in complement to existing supercomputing resources to run numerical simulations. In 2019, France was one of the earliest AI adopters, thanks to the French #AIForHumanity plan and 5 years later, the success is unconditional.

“Jean ZAY 4th generation has already proved in a few months that it is a wonderful machine to foster creativity at very high levels in very important and crucial fields of science and innovation. Installed in an unprecedented speed with the support of the teams of Eviden and IDRIS, this converged HPC & AI computing power reveals very important capabilities and capacities to help front-edge science teams in order to prepare great breakthroughs, e.g. in biology, health or astrophysics. As AI cannot wait it is now up to human genius to take, for the best, advantage of this artificial intelligence tool ! ” said Philippe Lavocat, GENCI chairman and CEO.

“The Jean Zay supercomputer is a key factor in France's attractiveness and contribution in terms of AI research at the highest international level. The French government has entrusted the CNRS with the responsibility of hosting and operating this research infrastructure, via its Institut du développement et des ressources en informatique scientifique (IDRIS), confirming its role as a key player in AI research. The exceptional capabilities provided by the Jean Zay 4th generation represent a tremendous opportunity for the entire French scientific and industrial community. It has already been proved in a few months.” said Antoine Petit, chairman and CEO of CNRS.

About

GENCI

Created by the French public authorities in 2007, GENCI (Grand Équipement National de Calcul Intensif) is a major research infrastructure. This public operator aims to democratise the use of digital simulation through high performance computing associated with the use of artificial intelligence, and quantum computing to support French scientific and industrial competitiveness.

GENCI is in charge of three missions:

- To implement the national strategy for the provision of high-performance computing resources, storage, massive data processing associated with Artificial Intelligence technologies and quantum computing, for the benefit of French

scientific research, in conjunction with the 3 national computing centres (CEA/TGCC, CNRS/IDRIS, France Universités/CINES).

- Supporting the creation of an integrated ecosystem on a national and European level
- Promoting digital simulation and supercomputing to academic research and industry

GENCI is a civil company 49% owned by the State represented by the Ministry in charge of Higher Education and Research, 20% by the CEA, 20% by the CNRS, 10% by the Universities represented by France Universités and 1% by Inria.

Regarding the national quantum strategy GENCI is partner together with CEA and Inria of HQI, the French HPC hybrid Quantum Initiative.

Follow GENCI on [LinkedIn](#), and visit their website <https://www.genci.fr/>

Follow HQI on [LinkedIn](#), and visit their website <https://www.hqi.fr/>

GENCI : contact@genci.fr +33(0)6.07.72.83.57

French National Centre for Scientific Research

A major player in basic research worldwide, the National Centre for Scientific Research (CNRS) is the only French organisation active in all scientific fields. Its unique position as a multi-specialist enables it to bring together all of the scientific disciplines in order to shed light on and understand the challenges of today's world, in connection with public and socio-economic stakeholders. Together, the different sciences contribute to sustainable progress that benefits society as a whole.