EXASCALE

Supplier selected!

EVIDEN has been selected to supply the new HPC system *Alice Recoque*. An announce formalized during SuperComputing.

GENCI

HPC at the service of knowledge

QUANTUM

International year

A busy and dimensioning one with a first quantum device now accessible, the launch of the Houses of Quantum network, etc.

ANNUAL REPORT 2025

TECHNOLOGICAL WATCH

Serving the Exascale

The Technological Watch Group assessed the potential breakthrough computing accelerators for AI and DNA-based memory storage technologies, etc.

AI

A one-stop-shop

AI Factory France is launched to support AI research and the development of business applications.

Sustainable HPC

Program at full speed

NumPEx had significantly expanded its scientific scope: new activities, robust engagement with the international community...

www.genci.fr

GENCI

In charge of **PROVIDING** high-performance computing and massive data processing resources, **GENCI'S MISSION**, at national and European level, is to promote the use of HPC combined with Artificial Intelligence and prototype quantum computing devices, for the benefit of scientific, academic and industrial **RESEARCH COMMUNITIES.**

More than 237 PFlop/s in 3 computing centers

- ***CINES* ADASTRA** Supercomputer 91,6 PFlop/s
  IDRIS Jean Zay Supercomputer 125,9 PFlop/s
  TGCC Joliot-Curie Supercomputer 20 PFlop/s
- For the entire IR*, 2489 research projects processed during the year, **with more than 4.6 billion core hours requested in CPU and 171 million GPU hours.**
- The three centers had more than **6,000 active users in 2025,** drawn from both academic research and private-sector research.
- **11 Thematic Committees**

** Research infrastructure (IR)*

FIVE shareholders

cea cnrs Inria France Universités MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR, DE LA RECHERCHE ET DE L'ESPACE

CONTENTS

The EuroHPC JU, together with the support of GENCI, is powering Europe's AI and supercomputing.

Anders Jensen,

EXECUTIVE DIRECTOR

The year 2025 marked a decisive step in consolidating Europe's high-performance computing (HPC) and advance computing technologies ecosystem, with the EuroHPC Joint Undertaking (EuroHPC JU) and GENCI working closely alongside numerous French and European partners.

The objective is clear: equip Europe with world-class, sovereign, and integrated computing capabilities that serve science, industry, and citizens. GENCI has actively collaborated with the EuroHPC JU on multiple initiatives to achieve this goal.

A major milestone was reached with the signing of the contract for the *Alice Recoque* exascale supercomputer, the second in Europe, to be operated by GENCI. This system will dramatically enhance our ability to tackle the most demanding scientific and industrial challenges, including artificial intelligence, climate modelling, and healthcare.

In parallel, 2025 saw tangible progress in the European hybrid HPC–quantum strategy. The inauguration of the *Ruby* quantum processor, alongside the installation of the *Lucy*, the photonic quantum computer, highlights the diversity and complementarity of the quantum technologies pursued by the EuroHPC JU with GENCI. Both systems are, very shortly, poised to provide user access and enable the exploration of workflows across materials science, meteorology, and engineering simulation, among others.

These advances are anchored in the ambitious EURO-QHPC integration project, coordinated by GENCI and funded by the EuroHPC JU, which brings together six quantum hosting entities and their associated supercomputers located across Europe. This unprecedented infrastructure will facilitate the coupling of multiple quantum technologies with leading HPC systems, accelerating the development of practical, real-world applications.

Looking ahead, the AI Factory France also promises to further strengthen the HPC–AI continuum, boosting Europe's competitiveness several other domains.

Collectively, these achievements reflect a shared commitment to excellence, openness, and European technological sovereignty, highlighting France's pivotal role, through GENCI and its partners, in shaping Europe's digital future.

The government has entrusted Inria with the role of a digital programs agency—focusing on algorithms, software, and applications—to strengthen France's digital sovereignty through research and innovation.

Its missions include developing a collaborative, forward-looking scientific vision; designing and managing programs that enable high-impact strategic partnerships; supporting the national economy through technology transfer; backing public policies; and coordinating European and international initiatives.

Its activities are structured around thematic programs that support national acceleration strategies and the "France 2030" initiative, whether in core digital fields such as artificial intelligence, post-exascale scientific computing, cybersecurity, network and cloud systems, software technologies for quantum computing, and virtual worlds, or in major application areas such as digital health and digital environmental solutions.

As part of the national AI strategy, GENCI—with support from Inria in its role as the government's coordinator for research initiatives—is leading the European project "AI Factory France (AI2F)," which serves as a national gateway to a suite of AI services accessible to businesses and academia, leveraging major national computing infrastructures.

Furthermore, the "Post-Exascale Scientific Computing" program includes initiatives in which GENCI is a key partner.

- The NumPEx PEPR (jointly led by Inria, the CEA, and the CNRS) aims to design and develop the software components that will power future exascale machines and to prepare key scientific and industrial application areas to take full advantage of these machines' capabilities.
- The France 2030-funded initiative "Components for AI," in collaboration with the CEA as part of its role as a program agency (ASIC), aims to develop modular acceleration components for AI and their software environment, relying on a joint hardware/software/application design to achieve its objectives of performance, reliability, and energy efficiency.
- The "Post-exascale" CSA, led by Inria, will bring together the European ecosystem, in close alignment with the national strategy.

GENCI is a key player in high-performance computing in France and Europe, playing a major role in developments related to AI.

Sophie Proust,

EXECUTIVE DIRECTOR OF THE PROGRAM AGENCY FOR DIGITAL SCIENCE AND TECHNOLOGY, INRIA

To develop impactful roadmaps, the ASIC Agency must rely on strong interactions with the French high-performance computing ecosystem. I would like to thank the GENCI teams for the quality of our exchanges since I took up my position last September.

Hughes Metras,

ASIC PROGRAM AGENCY EXECUTIVE DIRECTOR

The ASIC Program Agency is one of the seven agencies established in 2024 to structure research and development ecosystems in areas of major societal, economic, and sovereignty importance. In addition to its scientific monitoring and community-building missions, its primary role is to develop roadmaps and research programs aimed at delivering scientific breakthroughs that will align with industry technology and product roadmaps over a 5- to 10-year horizon. The ASIC Agency is responsible for the hardware dimension of digital technologies, in coordination with the Inria Digital Agency, which covers software, algorithms, and applications. In carrying out its missions, the ASIC Agency takes into account the entire value chain, starting from application challenges—particularly in digital infrastructures (computing and networks) and systems (transportation, robotics, and end-user devices)—and working upstream to components and materials.

The field of high-performance computing is undergoing rapid transformation, driven by the explosion of artificial intelligence and the imminent emergence of new computing paradigms such as quantum and analog computing. This evolution is a major driver of innovation in components and architectures. In this context, GENCI's leadership role as an operator of computing infrastructures for the scientific community is essential to maintaining in France a broad community at the forefront of advances in the field. This excellence enables the development of long-term visions for infrastructure requirements and, consequently, the creation of roadmaps for the components and systems that underpin them.

Building on the formalization of these needs, the two agencies have developed a research program on AI components (CAMELIA), with the ambition of delivering breakthrough AI accelerator solutions, particularly in terms of performance and energy efficiency. Jointly led by the two agencies, the program covers both the hardware and software dimensions of future heterogeneous system architectures based on chiplets. Finalized and approved in 2025, the program began in early 2026 and will run for approximately six years, mobilizing more than 20 research teams across France. In parallel, the ASIC Agency will lead a complementary program focused on the development of methodologies and design tools for heterogeneous systems-on-chip (SoCs), supporting the European design platform currently being launched. The objective is to foster the development of skills within a broader community of chip designers. GENCI teams will be invited to participate in the scientific and industrial steering committees of both programs.

Beyond the hardware/software/application co-design requirements for heterogeneous computing mentioned above, the roadmap development work conducted within the framework of the National Strategic Review has identified major innovation opportunities within computing systems. These include optical technologies for ultra-high-bandwidth, low-latency interconnections, advanced power conversion chains, and innovative architectures and components that converge data storage technologies (memory) with computing technologies (processors and accelerators). These topics will receive particular attention in the coming months from the Agency and its partners, with the aim of delivering high-impact innovations.

Interview

The strategic plan we are currently preparing will reinforce our position as a public operator ensuring strategic control over high-performance computing.

Mickaël KRAJECKI,

CEO OF GENCI

_ HOW WOULD YOU SUM UP THE YEAR 2025?

M. K. : Having taken up my responsabilities as CEO of GENCI in February 2026, I would like to pay tribute to my predecessor, Philippe Lavocat, for his work over the past ten years. I know GENCI well, having until very recently served as chair of the thematic committee on artificial intelligence. Philippe has handed over the reins of this remarkable research infrastructure to me, an organisation dedicated to putting digital excellence at the service of science and innovation. In the three main technological fields for which we are responsible – high-performance computing, quantum computing and AI – 2025 has been marked by decisive achievements: the selection by EuroHPC and the Jules Verne consortium of Bull as the supplier for the Exascale supercomputer *Alice Recoque*; the inauguration of *Jean Zay 4*; the commissioning of the *Ruby* quantum computer; and the launch of the AI Factory France project.

_ WHAT ARE GENCI'S MAIN CHALLENGES AND ISSUES FOR 2026?

M. K. : The first of these will lie in striking a balance between national and European strategies. This is due in particular to the expanding scope of EuroHPC and the emergence of AI Gigafactories, against a backdrop of particularly tense international relations.
The deployment and adoption of quantum technologies by the scientific community represents a second major challenge. This concerns not only the systems already installed at the TGCC, *Ruby* and Lucy, but also the system based on cat qubits, which we will be acquiring this year.
Finally, defining GENCI's strategy for the coming years is a major challenge. This is the purpose of the strategic plan we are currently preparing. It will reinforce our position as a public operator ensuring strategic control over high-performance computing.
At its heart lies our mission to provide researchers with access to computing infrastructure of the highest European and international standard. This will lead, in particular, to a strengthening of the complementarity between the three centres.

_ WHAT WORD DO YOU THINK BEST DEFINES THE YEARS AHEAD?

M. K. : Without a doubt, the word 'synergy'. It captures both the spirit and the essence of our day-to-day work in delivering large-scale European projects, as exemplified by *Alice Recoque*, as well as the spirit of mutual support that underpins relationships among GENCI staff, both internally and with external partners.

FOREWORD BY OUR SHAREHOLDERS

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR, DE LA RECHERCHE ET DE L'ESPACE

Jean-Luc MOULLET
General Director for Research and Innovation

cnrs

Antoine PETIT
Chairman and Chief Executive Officer

The year 2025 confirmed the central role of GENCI in the French and European high-performance computing (HPC) and artificial intelligence landscape. The success of the *Jean Zay* supercomputer, praised by both the scientific and industrial communities, clearly demonstrates the relevance of the strategic choices made in recent years.

Among the major achievements of 2025, I would highlight the selection and launch of the AI Factory France project under the EuroHPC program, which positions France as a key player in AI, not only for academic research but also for the many start-ups that benefit from our public computing resources. Another historic milestone was the signing of the purchase contract for the exascale supercomputer ***Alice Recoque*, which places our country at the forefront of high-performance computing. *Alice Recoque* will enable us to tackle major scientific and technological challenges while strengthening European sovereignty in a field where international competition is fierce. We should also note the deployment of two quantum computers, *Jade* and *Ruby*,** within HPC infrastructures, representing a significant step forward in the transformation of quantum technologies. Finally, the expansion of the ADASTRA supercomputer's capabilities at CINES will meet the growing and sustained demand for artificial intelligence applications.

These numerous examples illustrate the remarkable work carried out by the GENCI teams and the computing centers of its partners, who implement the supercomputers in the computing centers. Their expertise and commitment contribute to our collective success and enhance our ability to innovate together.

The year 2026 promises to be full of opportunities. We will pay particular attention to the assembly of the *Alice Recoque* supercomputer, whose commissioning in 2027 will mark a new milestone for France and Europe. Furthermore, the progress made in the field of quantum computing should enable us to expand GENCI's offerings and strengthen its position in this very promising area, for which it is essential to continue developing use cases.

In 2026, GENCI will therefore continue to play this key role in addressing the scientific and technological challenges of tomorrow.

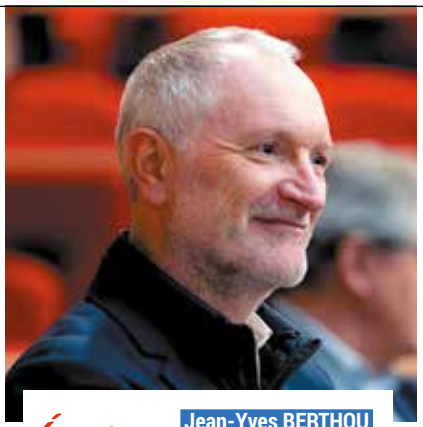
The success of the *Jean Zay* supercomputer, whose computing power was increased to 126 Pflops in 2024, remains unabated, and its adoption by research communities—particularly in artificial intelligence—has exceeded all expectations, with resource requests now approaching five times its capacity.

The journey does not end here. It is now essential to prepare for the replacement of the machine's oldest partitions, installed in 2019, in what remains a highly uncertain budgetary context. It is also necessary to continue and intensify support for research teams in their adoption and use of AI tools. Thus, 2025 marks the strengthening of the CNRS AI engineering team, which provides technical expertise and operational support for deploying projects on *Jean Zay*.

Moreover, the rapid evolution of the high-performance computing landscape, European initiatives

(such as the future exascale machine *Alice Recoque* at the TGCC), and the development of public-private partnerships (AI Factory France, future "gigafactories," etc.) require a rethinking of GENCI's positioning, role, and missions. **As a powerful tool for dialogue, resource sharing, and rationalization, GENCI enables collective responses to the ever-growing needs of our scientific communities.**

Finally, tribute must be paid to the memory of Ms. Hélène Mouchard-Zay, the youngest daughter of *Jean Zay*, who recently passed away. Those who had the privilege of meeting her and listening to her remarks during the dedication of our supercomputer to her father will forever remember a special moment with an exceptional woman.



Jean-Yves BERTHOU
Deputy Chief Executive Officer for Site Policy

Last year, we announced our plan to respond to the AI Factory call for proposals launched by EuroHPC. Co-led by GENCI and Inria—the coordinator of the National AI Research Strategy (SNRIA)—the AIF France project was approved for funding in March 2025. It brings together French computing resources provided by GENCI and all SNRIA stakeholders, foremost among them the nine AI clusters. **More than €800 million in projects funded by the SNRIA will contribute to the AIF France service offering, benefiting twelve industry sectors,** dincluding robotics, space and aerospace, energy, defense, healthcare, and agriculture.

The growth in requests for access to GENCI's AI computing resources observed in recent years is accelerating, accounting for 76% of GPU computing hours consumed. Nearly 60% of the projects submitted in 2025 in the field of AI are new projects, demonstrating the vitality of the field. 88% of the projects submitted by Inria use or develop AI technologies. These projects consumed 12% of GENCI's GPU capacity and originated from 9 of the University's 10 Inria centers, including Inria Chile.

For example, the project led by Pierre-Yves Oudeyer at the Inria center at the University of Bordeaux—which has been allocated 2 million GPU hours on the *Jean Zay* and Adastra machines—aims to create autonomous learning agents capable of exploring complex environments such as Minecraft or robotics and collaborating with humans via natural language. One application of this project is the development of models capable of tracking and predicting students' learning trajectories in order to offer personalized exercise plans.

Jean-Christophe Lombardo of the Inria center at the Université Côte d'Azur, meanwhile, is utilizing Genci's GPU capabilities as part of Pl@ntNet, a global participatory platform dedicated to plant identification and the monitoring of plant biodiversity, used by 20 million people worldwide. Research conducted using Genci resources has focused in particular on improving classification models and mapping species distribution on a large scale by cross-referencing citizen observations with satellite imagery (Sentinel) and environmental data.

FOREWORD BY OUR SHAREHOLDERS

Pierre-Alain MULLER
President at University of Haute-Alsace, digital referent for france universités

2025 was a pivotal year for high-performance computing and AI in France, with GENCI leading major breakthroughs. In May, the fourth extension of the *Jean Zay* supercomputer, equipped with 1,456 NVIDIA H100 GPUs, was inaugurated, quadrupling its power to 125.9 petaflops. Hosted by IDRIS (CNRS), *Jean Zay* is now Europe's flagship supercomputer for AI-driven research, supporting over 1,700 scientific projects in fields such as biomedicine, astronomy, and industry.

Meanwhile, the launch of the AI Factory France project, selected by EuroHPC, embodies a national and European ambition. Coordinated by GENCI and Inria, this unique hub brings together 20 public and private partners, including the 9 IA Clusters, France Universités, and CEA, to address challenges in computing, data, and AI skills. It provides infrastructure, training, and expertise services to the entire scientific and economic community, from startups to large corporations.

Finally, the convergence between numerical simulation and AI, initiated in 2019, is bearing fruit: for the first time in 2025, GENCI allocated more computing hours to "AI for Science" than to traditional AI, validating an integrated and innovative approach.

These achievements position France at the forefront of high-performance computing and AI, serving science and innovation.

Our FIVE shareholders

Marie-Hélène MATHON
Head of research infrastructures

The year 2025 marks a major milestone for GENCI and for the French scientific computing ecosystem. The gradual transition into the exascale era, the convergence of high-performance computing (HPC), artificial intelligence and quantum technologies, together with the strengthening of European initiatives, further reinforce GENCI's central role in both the French and European strategic landscape.

In this context, CEA, through the TGCC, is fully committed. The teams from GENCI, CEA and EuroHPC successfully conducted the procurement process that led to EuroHPC awarding Bull the contract for the future exascale supercomputer, *Alice Recoque*. Preparations at the TGCC continued with major infrastructure upgrades, carried out without disrupting scientific production and while meeting an exceptionally demanding schedule. This achievement reflects the commitment and expertise of the TGCC teams.

At the same time, a sustained effort is underway to optimize and port scientific applications to GPU-based architectures in order to support the growing adoption of hybrid HPC–AI workloads. These developments will enable increasingly complex simulations and the efficient processing of ever-growing volumes of data.

At the same time, the scientific and industrial communities are making increasing use of AI and HPC, driven by advances in foundation models and the growing need to process and analyze massive datasets. This convergence is transforming research practices and accelerating innovation across a wide range of disciplines. **In this rapidly evolving landscape, GENCI continues to support the scientific community by providing access to world-class computing resources, fostering the development of optimized software and applications,** and contributing to the emergence of a robust European digital ecosystem. Through its close collaboration with its academic, industrial and institutional partners, GENCI is helping to strengthen Europe's technological sovereignty and competitiveness while addressing the major scientific, economic and societal challenges of tomorrow.

France Universités

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR, DE LA RECHERCHE ET DE L'ESPACE
Liberté Égalité Fraternité

More about GENCI's shareholders

WHAT THEY SAY ABOUT US

→ In the press - On the Web

Les Echos

November 19, 2025

At its plant in Angers, Eviden — a company within the Atos Group — will manufacture France's first Exascale supercomputer: *Alice Recoque*. A project budgeted at €554 million.

At its plant in Maine-et-Loire, the products division of the Atos Group is preparing to begin production of France's first French exascale supercomputer. A machine that will have the capacity to exceed one exaflop per second—that is, one billion billion calculations per second—fifty times the capacity of the previous system. The project is estimated to cost €554 million over five years of operation. It is co-funded by EuroHPC, with a budget from the Digital Europe Program (DEP), and **by the Jules Verne consortium, led by France through GENCI and the CEA,** with participation from the Netherlands through Surf, and from Greece through GRNET.

L'USINEDIGITALE

November 18, 2025

Eviden, AMD, and SiPearl have been selected for *Alice Recoque*, France's new Exascale supercomputer.

France's next major supercomputer is moving forward. **On November 18, EuroHPC and the Jules Verne consortium (comprising GENCI, the CEA, and the Dutch and Greek organizations SURF and GRNET) announced the selection of Eviden** to build *Alice Recoque*, named after the French scientist of the same name, one of the first women to have made a name for herself in the field of artificial intelligence.

franceinfo:

November 29, 2025

Welcome to *Alice Recoque*, France's first Exascale supercomputer.

France is putting into operation its first supercomputer capable of performing one billion billion operations per second. **In addition to its scientific applications, it is intended to enhance France's technological appeal.** Alice Recoque was first and foremost a brilliant computer scientist and French artificial intelligence specialist, born in 1929 and who passed away in January 2021. She was a pioneer not only in technical research but also in addressing societal challenges.

radiofrance

May 14, 2025

The computing power of the *Jean Zay* supercomputer has just quadrupled.

One of the most powerful computers in France was officially unveiled on Tuesday in Orsay, in the Paris region. **It has been equipped with a fourth expansion module that enables it to perform nearly 126 million million million operations per second.** When we talk about a supercomputer, imagine a room the size of a tennis court filled with rows of large metal cabinets...

L'USINENOUVELLE

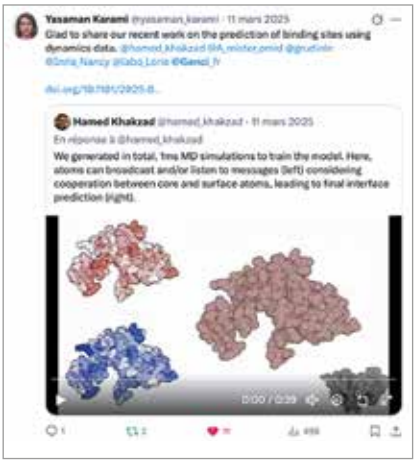
May 8, 2025

What will be the purpose of France's AI Factory, supported by the European Union?

The European Union has announced the launch of 13 AI factories, including one in France. This initiative aims **to support the development of computing power, with a range of tailored services, explains GENCI,** which will be in charge of coordinating the project.

WHAT THEY SAY ABOUT US

→ In the press – On the Web



→ IDRIS-*JEAN ZAY*

AFTER THE 4TH EXPANSION, RAPIDLY GROWING DEMANDS

Pierre-François LAVALLÉE
Director of IDRIS
((CNRS's center hosting the *Jean Zay* Supercomputer)

The national *Jean Zay* supercomputing infrastructure reached a major milestone in 2025 with the official inauguration of its fourth expansion, followed by its full deployment to the scientific community. This expansion represents a significant enhancement of France's national computing resources, directly addressing the growing demands driven by Artificial Intelligence (AI) and its exponential adoption across all scientific fields (AI4Science). The new capabilities have generated exceptional interest, with user demand two to three times higher than available capacity, underscoring their pivotal role in advancing research.

→ DALIA Supercomputer—an NVL72 rack consisting of 18 nodes, each with 2x NVIDIA Superchip GB200 (1x NVIDIA Grace processor with 72 Arm Neoverse v2 cores, 480 GB of LPDDR5X memory, and 2 NVIDIA Blackwell GPUs B200, each with 192 GB of HBM3e memory) connected to each other via the NVSwitch (130 TB/s of bandwidth, 5,000 coaxial cables for a total length of 3.2 km).

Supporting scientific communities remains a top priority in this context. The IDRIS offers **a comprehensive training program covering both HPC and AI (over 250 engineers and researchers trained in 2025). The institute also hosted its annual international hackathon in collaboration with Nvidia, supporting 10 innovative projects in 2025.** Additionally, IDRIS plays an active role in European initiatives, including the MINERVA project (minerva4ai.eu), focused on large-scale AI support, and AI2F (aifactory-france.eu). As a co-organizer of the « *Journée Deep Learning pour la Science* », IDRIS fosters knowledge sharing and collaboration on deep learning applications in scientific research.

Preparing for the future also involves proactive technological watch, open to the broader community. IDRIS hosts DALIA, the first Nvidia NVL72 Grace Blackwell architecture in Europe dedicated to academic research—a groundbreaking system that foreshadows the next generation of computing and inference architectures in the years ahead.

→ TGCC – *JOLIOT-CURIE*

QUANTUM DEVELOPMENT WAITING FOR... *ALICE RECOQUE*

Nicolas LARDJANE
TGCC's manager
((CEA's center hosting the *Joliot-Curie* Supercomputer)

In 2025, the Very Large Computing Centre of CEA (TGCC) continued the work started with GENCI on the development of quantum computing and preparations for Exascale computing.

The year was marked by a major breakthrough, with the deployment of the Ruby quantum simulator supplied by Pasqal. French and European researchers now have access to a 100-qubit machine, integrated into the TGCC environment. End 2025, the Quandela company has installed a 12-qubit quantum computer at the TGCC, based on a photonic approach, and a call for tenders is currently underway for a third technology, due to arrive in 2027. This equipment forms part of an exploration strategy led by the national HQI project, aimed at establishing a hybrid computing platform accessible to academic and industrial communities. Numerous training courses and initiatives across the regions have been launched, and initial feedback is promising.

At the same time, by the end of 2025, the TGCC had completed the main infrastructure upgrade required to accommodate the *Alice Recoque* Exascale supercomputer. This major evolution involved, in particular, increasing the electrical capacity (+24 MW), installing a 20 MW warm-water cooling loop, and adapting the computer rooms to meet the weight, power supply and space requirements of the new-generation computing cabinets. Following a tender process led by EuroHPC, Bull will manufacture and install the *Alice Recoque* machine at the TGCC in 2026 and 2027. With over 100 computing racks, *Alice Recoque* will offer the scientific community a computing capacity exceeding one exaflops thanks to an AMD GPU partition that is particularly effective in HPC and AI. Scalar computing capacity is also being integrated to eventually replace the *Joliot-Curie* supercomputer. I would like to highlight the unity of the three national supercomputing centres and GENCI in this project, particularly with the establishment of an advanced application support team from 2026 onwards to assist users.

Beyond the technical aspects, these developments reflect the determination of France and Europe to remain key players in the broader high-performance computing ecosystem. By combining high-performance computing, artificial intelligence and quantum computing, the TGCC and GENCI are strengthening their capacity to support some of the most ambitious scientific projects, whilst anticipating the future needs of the scientific communities.

→ The logo for the future *Alice Recoque* Supercomputer was revealed in 2025.

→ CINES-ADASTRA

OUTLOOK FOR AN EXTENSION CONFIRMED

Michel LABADIE
Head of CINES
(France Universités's center hosting the ADASTRA Supercomputer)

In 2025, CINES will continue to play its pivotal role in high-performance computing and supporting scientific applications.

With a stable computing power of 90 Pflop/s, the ADASTRA supercomputer reaffirms its central role by supporting more than 1,200 users. The platform maintains a high level of service, with availability exceeding 99% for CPUs and 98% for GPUs, as well as robust user support.

To make it easier to get started with the system, CINES has also expanded its support services by launching "quick start" instructional videos. Available on Canal U and YouTube, these videos allow users to familiarize themselves more quickly with the ADASTRA environment and its best practices, while optimizing the time spent by support teams.

CINES is also playing a key role in preparing researchers for the future exascale supercomputer, Alice Recoque. The expertise of the CINES teams and the ADASTRA architecture constitute a major asset in supporting scientific communities as they adapt their code to run on AMD technology, in collaboration with GENCI, the CEA, and IDRIS.

As part of this initiative, CINES has continued its commitment to the scientific communities by organizing hackathons dedicated to optimizing code on GPU architectures, in partnership with HPE and AMD. This momentum will continue in 2026 with several events already scheduled. Dialogue with users continued

→ User training provided at CINES in 2025 by the support teams.

CINES staff taken in front of the computing center.

throughout the year, notably through the holding of four Computing User Committees. These meetings provided an opportunity to share platform updates, gather feedback on community needs, and fuel a process of continuous improvement of the services offered.

The year 2025 saw a strengthening of activities related to artificial intelligence, with the establishment of a dedicated division and involvement in European projects such as Minerva, contributing to supporting new applications and fostering scientific innovation.

Finally, in late 2025, GENCI confirmed the expansion of the ADASTRA supercomputer, marking a new milestone in strengthening national high-performance computing capabilities. This expansion involves the addition of 84 nodes equipped with MI300 processors, increasing the number of APUs from 112 to 448 and reaching a total capacity of more than 2,000 accelerators (GPUs and APUs). The storage infrastructure will also be enhanced with the addition of 3 Po of high-throughput, low- latency space in the /scratch area, in order to support the growing demand for computing resources and the increasing volumes of data processed, particularly in the field of artificial intelligence.

A YEAR'S DIARY

CONTENTS


80
2
1
2
3

→ Exascale

Selection of supplier for *Alice Recoque* supercomputer

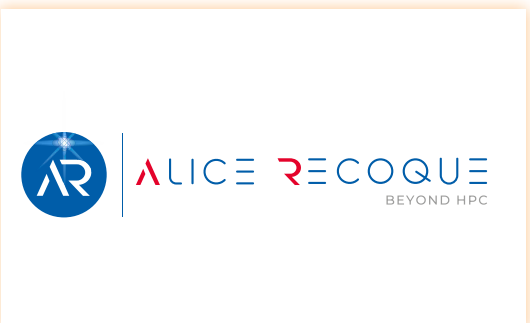
EuroHPC and the Jules Verne consortium announced the selection of EVIDEN to supply the *Alice Recoque* supercomputer on November 18, 2025, during Super Computing 2025.

The year 2025 was mainly devoted to selecting the supplier for the *Alice Recoque* computer. This procurement procedure, led by EuroHPC, began in September 2024 and was completed in November 2025.
GENCI drafted the specifications in collaboration with the CEA. Five dialogue sessions were organized with the candidates between January and June 2025. These sessions enabled the specifications to be developed into the final version published in July 2025. After the offers were evaluated by a committee of three experts, the selection of EVIDEN, now Bull, to supply the *Alice Recoque* supercomputer was announced jointly by EuroHPC and the Jules Verne consortium on November 18, 2025.

> **Eviden's selection to supply the *Alice Recoque* supercomputer marks an important milestone in France's Exascale project.**

CORINNE BEAL, EXASCALE PROJECT MANAGER AT GENCI

→ The logo for the *Alice Recoque* supercomputer highlights both the humanistic values inherent in science—its universalism—within the context of French and European sovereignty, accompanied by a tagline that conveys the ambition behind this machine.



A NEW GENERATION OF SUPERCOMPUTER

The *Alice Recoque* supercomputer will include a unified partition composed of next-generation AMD processors: AMD EPYC "Venice" CPUs and AMD Instinct MI430X GPUs interconnected with a Bull network solution (BXIv3) and integrated into a BullSequana XH3500 platform.
For scalar workloads only, *Alice Recoque* will integrate a dedicated scalar partition with more than 100,000 cores using European SiPearl RHEA2 128-core ARM-based processors. Pre/post-processing or heavy SMP workloads will be handled by an additional partition. Comprising 94 racks, the system will have the capacity to reach the exaflop threshold with 25% fewer racks and components than other exascale systems. *Alice Recoque* will integrate new European technologies into processors/accelerators, networks, storage, cooling, and infrastructure, as well as HPC/AI/QC middleware and

Announcement of the selection of Bull/EVIDEN as supplier for *Alice Recoque* supercomputer at SC25.

end-user applications.
The computer will be installed at the CEA's Très Grand Centre de Calcul (TGCC) in the last quarter of 2026 and will go into production at the end of 2027.

JULES VERNE CONSORTIUM

Greece, represented by GRNET, joined the Jules Verne Consortium in July 2025. We also continued discussions with academic and industrial partners who may join the consortium. Following the signing of the hosting agreement in June 2024, we signed with EuroHPC the Administrative Agreement, which governs the payment of contributions by consortium member countries toward acquisition costs, and the Grant Agreement, which defines the terms under which EuroHPC will disburse funding for operating costs.

More about *Alice Recoque* Supercomputer

Unveiling of the *Alice Recoque* supercomputer logo

This logo consists of several elements. The blue sphere with the initials "AR" refers both to the atom or elementary particle and to the globe. The star adorning this sphere symbolizes the compass, like the one that guided sailors, and also embodies the "promise" of new discoveries. **This refers to *Alice Recoque*, a pioneer in AI, as well as to the machine itself,** which holds collective, scientific, societal, and therefore universal promise.

→ Artificial Intelligence

AI Factory France, the one-stop-shop fostering AI in France

Launched in late 2025 AI Factory France is a one-stop shop providing computing resources, expertise, and services to support AI research and the development of business applications.

Initiated by the European Commission as part of the AI Continent strategy and selected by EuroHPC in March 2025, AI Factory France (AI2F) marks a key milestone for European digital sovereignty, strengthening the continent's strategic autonomy and competitiveness in the field of artificial intelligence (AI).

A SOVEREIGN ECOSYSTEM FOR EXCELLENCE IN AI

Coordinated by GENCI in partnership with Inria under the Digital Programs Agency, the project has a €30 million budget over three years, dedicated to the development of high-impact infrastructures, services, and initiatives. It brings together 16 academic and innovation partners, including CNRS, CEA, AMIAD, and the universities leading the AI Clusters, helping to structure a coherent and interconnected national ecosystem. Integrated into a European network of 19 AI Factories and 13 hubs, AI2F works closely with its counterparts, notably in Ireland, to pool resources, share expertise, and strengthen European momentum.

→ AI2F work package leaders gathered at the project kick-off event at Station F **From left to right:** Arnaud Renard, Laurent Patoux, Raphaëlle Achach, Cedric Auliac, Maike Gilliot.

More about AI Factory France

Announcement of the collaboration between NVIDIA and AI Factory France by Howard Wright (VP Startups, NVIDIA) at an event at Station F.

Accelerating AI innovation

AI Factory France accelerates AI adoption among startups and SMEs through a unique partnership with the NVIDIA Inception and Connect programs. **This initiative provides access to advanced computing infrastructure and specialized AI services.** Building on GENCI's experience (over 80 startups supported and 1,700 projects per year), AI2F strengthens innovation and competitiveness at the European level.

SERVING RESEARCH AND INDUSTRY

AI2F aims to expand access to computing resources, notably through the exascale supercomputer *Alice Recoque*, a key lever for developing and training next-generation AI models, while structuring a public-private continuum aligned with industrial needs.

The project facilitates the use of data and models by improving accessibility, promotes partner-developed tools, and provides expert support in AI and data science across the full project lifecycle. It also supports talent development, training, and mobilizes venture capital to accelerate innovation. The objective is to deploy these services at the European level to enable the widespread adoption of AI within a competitive and responsible framework, notably through sector-based structuring. AI2F fosters communities across 12 key sectors such as robotics, defense, and healthcare, encouraging synergies between research, innovation, and industry.

Launched at Station F in late 2025, its rollout is accelerating through European collaborations, positioning this initiative as a cornerstone of Europe's response to AI-related challenges.

> **AI2F provides an essential response to the needs of both the research and business communities in AI, at national and European levels.**

CÉDRIC AULIAC,
AI PROGRAM MANAGER AT GENCI

➜ Artificial Intelligence

Jean Zay 4: AI accelerates French research

With *Jean Zay 4*, GENCI and IDRIS have provided French scientists with unmatched power in Europe: 126 petaflops, 1 456 H100 GPUs, and over 3.8 million GPU hours mobilized. *The Grands Challenges*, the qualification phase for each new computing platform, enabled ten projects to push the boundaries of simulation and AI in strategic fields.

A Supercomputer Serving Scientific Sovereignty! In December 2025, *Jean Zay 4* became fully operational: 126 Flops peak, optimized cooling, waste heat reuse, and energy-efficient AI training. The result of an exemplary competitive dialogue between GENCI, IDRIS, Eviden, NVIDIA, and DDN, it positions France at the heart of Europe's high-performance computing and responsible AI ecosystem.

GRANDS CHALLENGES: QUALIFICATION AND BREAKTHROUGHS

Since 2008, each new GENCI supercomputer opens with a "Grands Challenges" phase: a production qualification period where the system is tested at extreme scale on reference codes and major scientific projects.
Few projects are selected—they mobilize a very significant share of resources—and are chosen for their major scientific or societal impact and their ability to push the machine to its limits. This phase validates system robustness and stability under extreme conditions while granting pioneering teams—accepting potential instabilities—a temporary massive access to unprecedented power, enabling breakthroughs otherwise impossible. The *Jean Zay 4* Grands Challenges (2024–2025) mobilized over 3.8 million H100 GPU hours across ten landmark projects in strategic domains: astronomy, biology, quantum chemistry, health, autonomous driving, and sovereign language models.

Cover of the Grands Challenges hightlighting projects developed using *Jean Zay 4*.

OVERVIEW OF GRANDS CHALLENGES ON *JEAN ZAY 4*

- **AI at the heart of astronomy**

AION-1, by the CNRS: unifies multimodal analysis of astronomical data (images, spectra, physical measurements) in a 3-billion-parameter foundation model trained on 120 TB. It predicts physical properties with unprecedented accuracy and adapts to new tasks in just hours.

- **Molecular biology and its dynamics**

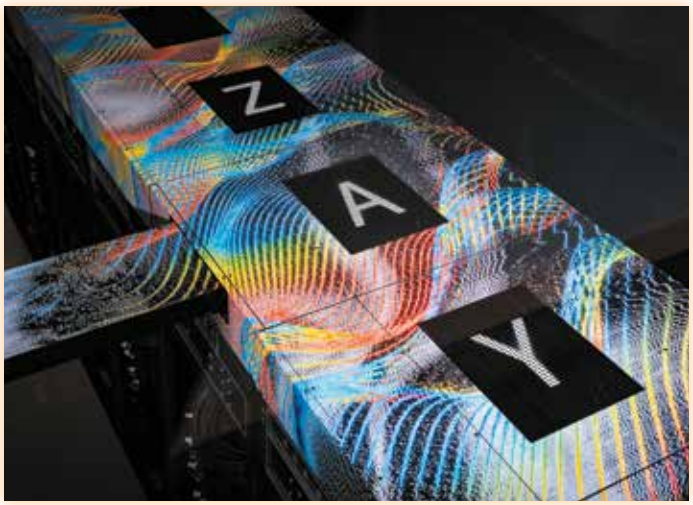
Ion Channels and Genetic Mutations, by THE CNRS: By building an atlas of activation kinetics, this work supports the development of precision therapies, classifying mutations according to their functional effects or guiding drug design.
DynaRepo, by University of Nancy: Captures over 1.1 millisecond of molecular dynamics across 700 systems, providing France's first repository of macromolecular conformations. It now fuels AI models sensitive to the movements of living systems.

- **Quantum chemistry and drug design**

FeNNix-Bio1 by the startup Qubit Pharmaceuticals: First foundation model for reactive atomistic simulations

> ***Jean Zay 4* transforms computing power into an accelerator of knowledge and innovation through AI.**

PHILIPPE LAVOCAT,
CEO OF GENCI



View of the compute racks of the 4th extension of the *Jean Zay* supercomputer: designed by the French startup *Obvious*.

in condensed phase with quantum accuracy, achieving hydration free energy errors of 0.37 kcal/mol on millions of atoms, simulates matter at atomic scale (89 elements) with DFT precision, accelerating materials and drug discovery

- **AI for health**

H-optimus-1 by the startup Bioptimus: becomes the histology reference: trained on 2 billion images, it outperforms existing feature extractors on 23 biomedical tasks.

- **AI for security**

Autonomous Driving and Vision, leaded by Valeo AI: learning from 1,800 hours of raw dashcam videos, achieving competitive performance on critical scenarios without annotations.

- **Sovereign and French-language models reducing carbon footprint and dependence on foreign clouds**

Lucie-7B, leaded by the startup Linagora: first multilingual LLM with balanced performance (capable of correcting biases), trained on open and traceable data.
Pleias 1.0: first family of language models trained exclusively on open data (Common Corpus).
Vigogne LLM: series of SLMs (Small Language Models) specifically optimized for French.

Controlled Societal and Environmental Impact

These projects combine extreme performance with sobriety: **low carbon footprint (nuclear electricity), waste heat recovery, and energy-optimized training.** They embody sovereign, ethical AI serving the common good—from precision medicine to energy transition.

For complete description of Grands Challenges on *Jean Zay 4*

→ Quantum

GENCI's international year of Quantum

While 2025 was declared *« International Year of Quantum »* by the United Nations, it was an especially busy and dimensioning one for GENCI in that field. The HQI platform is developing, with a first quantum device now accessible, the installation of the second one, and the emission of a tender for the third one. GENCI was also quite active in supporting and promoting French and European quantum ecosystems, through the launch of the Houses of Quantum network, and of course the *Quantum Zone* at *Vivatech* 2025!

This network was already able to form an international collaboration with QuantumBW, which supports the Baden-Württemberg ecosystem in Germany, signed on October 7th, 2025, at the Quantum Effects fair in Stuttgart. International collaborations were the main focus of a workshop organized by the Houses of Quantum on December 5th, 2025, at EDF Lab in Palaiseau, as a side event of the QUEST-IS conference, including contributions from Canada (PINQ2), Japan (RIKEN) and Germany (QuantumBW and the Munich Quantum Valley). This workshop will hopefully give rise to numerous discussions and collaborations between ecosystems in the coming months. Topic-based committees will be launched soon to foster specific exchanges between the different Houses of Quantum on algorithm-related topics, and coordinate new initiatives such as hackathons or even summer schools.

PROMOTING LOCAL HYBRID QUANTUM COMPUTING ECOSYSTEMS

This network was already able to form an international collaboration with QuantumBW, which supports the Baden-Württemberg ecosystem in Germany, signed on October 7th, 2025, at the Quantum Effects fair in Stuttgart. International collaborations were the main focus of a workshop organized by the Houses of Quantum on December 5th, 2025, at EDF Lab in Palaiseau, as a side event of the QUEST-IS conference, including contributions from Canada (PINQ2), Japan (RIKEN) and Germany (QuantumBW and the Munich Quantum Valley). This workshop will hopefully give rise to numerous discussions and collaborations between ecosystems in the coming months. Topic-based committees will be launched soon to foster specific exchanges between the different Houses of Quantum on algorithm-

→ Panel session on the *Quantum Zone* by HQI at *Vivatech* 2025 on the transition from fundamental research to industry, in the presence of Philippe Baptiste, minister of Higher Education, Research and Space.

The *Ruby* supercomputer (as *JADE*) represents a major advance in the implementation of quantum computing whithin the European scientific and industrial landscape.

The French-German inauguration of *Ruby* and *JADE*

On November 13, a triplex inauguration between TGCC, Brussels and FZJ was organized for the *Ruby* and *JADE* quantum computers, provided by Pasqal for the European project HPCQS. It was an opportunity to recall the motivations behind the acquisition of these two pilot systems and their integration in the European quantum strategy.

related topics, and coordinate new initiatives such as hackathons or even summer schools.

WELCOMING FIRST QUANTUM DEVICES IN THE HQI PLATFORM

The 100+-qubit analog quantum computer provided by Pasqal for the European project HPCQS was commissioned and inaugurated on November 13th, 2025, along with its twin system *JADE*, hosted at Forschungszentrum Juelich (FZJ), in Germany. *Ruby* is now available for the HQI platform end-users. After setting up a remote access to Quandela's technology for the platform end-users in March 2025, a 12-qubit photonic quantum computer called *Lucy*, acquired by EuroHPC Joint Undertaking to Quandela, was installed at TGCC starting October 15th, 2025. Two new tenders were launched in September and October 2025 to expose additional technologies within the platform: the first one, EuroSSQ-HPC, was emitted by EuroHPC to acquire a spin qubit device hosted by SURF in the Netherlands, and the second one was launched by GENCI to acquire a 18-cat-qubit system that will be hosted at TGCC. The latter will be delivered and installed early 2027. 2025 thus allowed the HQI platform end-users to start accessing actual quantum devices operating in a supercomputing center environment.

A YEAR'S DIARY

OFFICIAL KICK-OFF MEETING OF A CRUCIAL COLLABORATION FOR A EUROPEAN HYBRID QUANTUM COMPUTING PLATFORM

The EuroQHPC-Integration project kick-off meeting was held in Krakow, on the occasion of the EuroHPC Summit, on March 19th, 2025.
This project includes 30 partners from 17 European countries and is coordinated by GENCI. It aims at federating the HPC-QC integration strategies for six systems based on complementary hardware technologies, acquired by EuroHPC and hosted in six European supercomputing centers : PSNC, in Poland (EuroQCS-Poland), BSC-CNS, in Spain (EuroQCS-Spain), LRZ, in Germany (Euro-Q-Exa), CINECA, in Italy (EuroQCS-Italy), IT4Innovations, in Czech Republic (Lumi-Q), and GENCI/CEA, in France (EuroQCS-France).
The partners will work together to harmonize the user experience across the six platforms, foster expertise sharing between the support teams, set up joint applications and benchmark libraries, and draft joint feedback towards European and international normalization and standardization initiatives. First results should be available as soon as 2026.

Inauguration of the *Ruby* and *JADE* quantum computers: a major step towards a European hybrid HPC-QS infrastructure.

> **This inauguration is a major step towards the implementation of quantum computing in the European scientific and industrial landscapes.**

PHILIPPE LAVOCAT,
FORMER CEO AND CHAIRMAN OF THE BOARD OF GENCI

THE *QUANTUM ZONE* BY HQI AT *VIVATECH* 2025

HQI participated in the *Vivatech* conference in Paris for the fourth time this year, and it was particularly visible. The program organized a panel session including, amongst other guests, Mr. Philippe Baptiste, the French Minister of Research and Higher Education. Three other French ministers visited the booth on the General Public day. Canada was the country of the year and they sent a delegation to the *Quantum Zone* where they had a long discussion on possible partnership opportunities between the two ecosystems. 13 start-ups were hosted on the Quantum Zone for the four days of the event, and other programs from the French National Quantum Strategy, such as MetriQs-France, AtomQtrl and PEPR-Quantique, presented their activities. The booth was very much crowded through the whole event, which led the organizers to envisioning a wider surface, as well as a larger presence of the French National Quantum Strategy for the 2026 edition.

For more about *Ruby* and *JADE*

→ HPC sustainability

The Exascale *NumPEx* program at full speed in 2025

The year 2025 was highly productive for the *NumPEx* program, which significantly expanded its scientific scope. This growth was evident through the introduction of new activities and robust engagement with the international community.

NumPEx launched its first HPC/HPDA/AI call for projects at the end of 2024. With a €4 million budget, the call resulted **in the funding of four projects:**

- **DAIMOS:** a modular and energy-efficient software stack for large-scale deep learning, integrating optimized distributed training algorithms for graph neural networks and reusable HPC tools.
- **SAGE-HPC:** an open and scalable software platform for multi-fidelity optimization of complex physical problems, integrating Bayesian optimization, deep reinforcement learning, and AI-guided hybrid strategies.
- **Koktails:** contributes to an open-source software stack for GPU-based supercomputers through the Kokkos programming model, integrating AI and Python to ensure performance portability and facilitate the transition of legacy HPC applications to these new architectures.
- **ASTRA:** a unified approach to developing workflows for the analysis of massive scientific datasets deployable across federated infrastructures, using the *Square Kilometer Array* (SKA) as a use case.

→ The first *YoungPEx* meeting took place in November 2025 at the IRIT laboratory of Toulouse and was focused on three major subjects – sustainable HPC, the use of generative AI assistants in HPC research and equality/diversity in the HPC scientific community.

In collaboration with InPEx, *NumPEx* organized a workshop in Kanagawa, Japan, in April 2025, bringing together around one hundred European, American, and Japanese participants to discuss key challenges in HPC, including the growing integration of AI, workflows, the digital continuum, and co-design. Several workshops were held, including one entitled *"AI and HPC: Sharing AI-Centric Benchmarks of Hybrid Workflows"*, which focused on:

> ***NumPEx* is accelerating the integration of AI, HPC and HPDA through international collaborations and innovative projects in preparation for the Exascale supercomputer *Alice Recoque*.**

JÉRÔME BOBIN,
CO-DIRECTOR OF THE *NUMPEX* RESEARCH PROGRAM

A YEAR'S DIARY

- Building a shared understanding of AI-driven HPC/HPDA workflows;
- Exploring coupling models between AI and HPC/HPDA and identifying common execution patterns;
- Developing collaborative benchmarks, proxy applications, and validation metrics;
- Creating standardized tools to assess performance and scalability;
- Defining a roadmap with milestones for future events.

The next InPEx workshop will take place in April 2026 in Rio de Janeiro, where *NumPEx* will lead several sessions on post-exascale roadmapping, the use of generative AI for application design, and HPC-AI hybridization.

The *YoungPEx* network

In 2025 *NumPEx* launched its *YoungPEx* committee, **which is an interdisciplinary network that gathers young researchers (PhD, post-docs and engineers) from all the components of the *NumPEx* program.** The main objectives of this working group are to stimulate new research dynamics and exchanges between the various *NumPEx* projects, create synergies between researchers, and prepare for future collaborations.

➜ HPC sustainability

GENCI's Technological Watch Group serving the Exascale vision

The CVT has also strengthened its collaborations with disruptive startups. With Arago, it is exploring photonic accelerators for AI. With Biomemory, it is studying data storage using synthetic DNA. This technology promises up to 1,000 times less physical footprint than tape-based solutions, with exceptional stability for very long-term cold storage and an extremely low environmental impact. A first prototype is planned for 2027 to evaluate its performance and the new data duplication services it could enable.

The primary mission of the CVT is to identify and test emerging technologies in the fields of Exascale computing, artificial intelligence, processing of instrumental data, and storage. This monitoring effort relies on a network of experts from GENCI, the three national centers, regional stakeholders (MESONET), and HPC and AI application developers.

KELLY: AN ARM PROTOTYPE TO PREPARE FOR *ALICE RECOQUE*

In 2025, the CVT contributed to the sizing and acquisition of Kelly, a scalar prototype based on ARM architecture, installed at CEA DAM. Equipped with 8 dual-processor NVIDIA Grace servers, it replaces the A64FX partition of *Joliot-Curie* and will enable user communities to prepare for the Rhea2 (SiPearl) scalar partition planned for late 2027 on *Alice Recoque*.

PARTNERSHIPS WITH FRENCH STARTUPS

The CVT has also strengthened its collaborations with disruptive startups. With Arago, it is exploring photonic accelerators for AI. With Biomemory, it is studying data storage using synthetic DNA. This technology promises up to 1,000 times less physical footprint than tape-based solutions, with exceptional stability for very long-term cold storage and an extremely low environmental impact. A first

prototype is planned for 2027 to evaluate its performance and the new data duplication services it could enable.

STRATEGIC PRIORITIES 2026-2028

In response to geopolitical tensions and dependence on foreign suppliers, the CVT is intensifying its efforts: development of AI inference with a focus on all-optical solutions, strengthening collaborations with private and public stakeholders (notably the Ministry of the Armed Forces), and exploration of alternative technologies such as DNA storage. Despite a temporary slowdown linked to the Exascale workload, 2026 will be marked by a strong acceleration in forward-looking initiatives to anticipate the needs of *Jean Zay* Next, the AI Factory France, and to contribute to the emergence of European solutions.

Kelly, a scalar prototype based on the ARM architecture, deployed at CEA DAM.

“Anticipating technological breakthroughs is essential to preserving France's capacity to compute, innovate, and make sovereing decisions.”

ÉRIC BOYER,
CHIEF PROJECT OFFICER HPC/HPDA/AI AT GENCI

AT THE SERVICE OF RESEARCH

CONTENTS

HOW DOES GENCI WORK?

GENCI
HPC at the service of knowledge

SERVICES

- HARDWARE AND SOFTWARE TECHNOLOGICAL WATCH
- FREE ACCESS TO HPC AND HYBRID HPC/QUANTUM COMPUTING SYSTEMS FOR OPEN RESEARCH
- ACCESS TO STORAGE CAPACITIES (FROM FLASH TO LONG TERM STORAGE USING MAGNETIC TAPES)
- USER SUPPORT
- TRAINING
- DISSEMINATION OF SCIENTIFIC RESULTS

EuroHPC Joint Undertaking

FRANCE 2030

anr agence nationale de la recherche

SU

ADASTRA

HPE Cray EX4000 aggregated of 132,7 Pflop/s

3 partitions - AMD based :
Scalar, converged and APU architecture
- GENOA : Scalar 3,9 Pflops/s - 104 448 cores
- MI250X : Converged 74 Pflop/s - 1424 GPUs
- MI300A : APU 54,8 Pflop/s - 448 APUs

(Including 336 APUs extension - July 2026)

ALICE RECOQUE

ALICE RECOQUE
BEYOND HPC

EuroHPC Exascale system for addressing HPC and AI Factory at scale

- Bull Sequana XH3500 of > 1 EFlop/s sustained unified nodes using AMD Venice CPUs and MI430x GPUs
- scalar nodes using SiPEARL Rhea2 CPUs
- federated by Bull BXIv3 400 Gb/s interconnect

BULL/

Conver
- 36 NV
- 72 NV
- 17 TB
- 14 TB

QUA

RUBY
QUANTUM

A 100 qubit analogue quantum computer based on neutral atoms provided by Pasqal (HPCQS)

LUCY
QUANTUM

A 12 qubit digital quan computer based on ph provided by Quandela (EuroQCS-France)

UPERCOMPUTERS

JOLIOT CURIE

Bull Sequana XH2000 of 13.1 PFlop/s

Multiple partitions offer a wide variety of processors
Scalar nodes: 12 Pflop/s
• ROME: AMD – 2292 nodes
– More than 293000 cores
Converged nodes: 1,1 Pflop/s
– V100 – 32 GPUs

JEAN ZAY

HPE/BULL SGI 8600/XH3000 of 125,9 Pflop/s

4 partitions - scalar and converged
Scalar: 2,3 Pflop/s – Intel CSL: 720 nodes
– 28 800 cores
Converged nodes: 121 Pflop/s
• NVIDIA V100 – 15,5 Pflops/s – 1832 GPUs
• NVIDIA A100 – 8,2 Pflop/s – 416 GPUs
• NVIDIA H100 – 99,9 PFlop/s – 1456 GPUs

DALIA

NVIDIA GB200 NVL72 of 2,9 Pflop/s

ged architecture, composed of 36 nodes:
VIDIA Grace Arm Neoverse v2 – 72 cores
VIDIA Blackwell GPUs
of LPDDR5X
of HBM3E memory federated using NVLink

KELLY

BULL NIVIDIA Grace-Grace

Scalar architecture
• 8 NVIDIA Grace-Grace CPUs nodes
• 240 GB LPDDR per node
• Infiniband 400 Gb/s

PROJECTS

AI FACTORY FRANCE AI2F

A federated one-stop-shop of infrastructures and services for the AI community

EPICURE & MINERVA

A pan European task forces for User support in HPC and AI

AST ALICE RECOQUE

A tailored user support team for preparing the inception and supporting users on Alice Recoque

EPI2 / EUPEX

Developpement of EU microprocessor technologies

NTUM COMPUTERS

tum
otonics

KITY QUANTUM

A 18 qubit early FTQC quantum computer based on cat qubits provided by Alice & Bob

EuroSSQ-HPC

SURF

Remote access to a 16 qubit quantum computer based on spins, acquired by EuroHPC JU and hosted by SURF (NL)

EUROQHPC-I

EuroQHPC-I

EuroHPC QC Hosting Entities toward HPC and QC integration

TERMS OF ACCESS TO COMPUTING RESOURCES

With the arrival of the H100 partition on *Jean Zay*, a new type of access is available for large-scale projects with significant societal impact. As the selection process for these projects is not yet finalized, we have selected 6 projects from the Regular Allocations to test the Strategic Allocation model: a very large amount of computing resources (hundreds of thousands of accelerated computing hours) over a short period (two or three months).

National computing and storage resources are made available free of charge through one-year renewable allocations for scientific projects in high-performance computing (HPC), artificial intelligence (AI), and quantum computing, originating from both public and private sectors (with a SIRET number). To be eligible, open research projects must result in one or more publications at the end of the allocation period. Two steps are required: first, submitting a resource request for one or more supercomputers and obtaining approval, then applying for an account on the approved supercomputers. Both requests are processed via the portal www.edari.fr for the three national computing centers (CINES, IDRIS, and TGCC). In total, 14% of applications come from industry, including 9% directly from private entities such as startups.

TYPES OF ALLOCATIONS

In 2024, a new allocation type—Strategic Allocations —was added alongside Regular Allocations and Dynamic Allocations . Regular Allocations are open twice a year to project leaders requiring substantial computing resources: over 50 million CPU core-hours or 50,000 GPU hours. They require scientific and technical evaluation before approval. Dynamic Allocations , which are easier to obtain, are available year-round and are intended for less resource-intensive projects. They are also accessible to PhD students and Master's 2 students, whereas Regular Allocations are strictly reserved for permanent research staff (teachers, researchers, engineers, postdocs). The entire GENCI resource allocation process remains ISO 9001 certified.

Strategic Allocations

With the arrival of the H100 partition on *Jean Zay*, a new type of access is available for large-scale projects with significant societal impact. As the selection process for these projects is not yet finalized, we have selected 6 projects from the Regular Allocations to test the Strategic Allocation model: a very large amount of computing resources (hundreds of thousands of accelerated computing hours) over a short period (two or three months).

Explore the access terms infographic

THE COMPUTING RESOURCES AVAILABLE IN 2025

GENCI's computing capabilities (in Pflop/s) are provided over 90% by partitions equipped with GPU accelerators. We also have significant CPU partitions and, since this year, QPU (quantum computing) partitions.

ADASTRA

The CINES hosts a supercomputer from manufacturer HPE with a power of 91 PFlop/s, featuring a scalar partition of 544 nodes composed of AMD EPYC Genoa representing 104,448 cores and two converged partitions, one of 356 nodes composed of AMD MI250x totaling 1,424 GPUs and one of 28 nodes composed of AMD MI300A totaling 112 APUs.

Discover details of calculators on eDARI

JOLIOT-CURIE

The TGCC hosts a supercomputer from manufacturer Atos with a peak power of 13 PFlop/s, composed of 3 production partitions including 2 scalar partitions, one ARM prototype partition and a quantum platform. A first scalar partition of 2,292 AMD Rome nodes for 12 PFlop/s and 293,376 cores. The second scalar partition based on Intel Skylake was permanently shut down in November. Finally, to the accelerated partition of 32 Nvidia V100 nodes for 1.2 PFlop/s and 128 GPUs is added a prototype partition composed of 8 nodes equipped with two scalar processors Nvidia ARM Grace 72 cores, aiming to prepare codes for the transition to the European Exascale machine. The Quantum platform consists of a quantum emulator QLM from Atos and a physical quantum computer with cold atoms for the computer Pasqal installed in the autumn.

JEAN ZAY

The IDRIS operates a supercomputer with a peak power of 126 PFlop/s composed of a scalar partition and 3 converged partitions. A first scalar partition of 720 Intel Cascade Lake nodes for 2.5 PFlop/s and 28,800 cores. An accelerated partition of 15.5 PFlop/s composed of 437 accelerated nodes based on Nvidia V100 GPUs totaling 1,832 GPUs, another accelerated partition of 8.2 PFlop/s composed of 52 accelerated nodes based on Nvidia A100 GPUs totaling 416 GPUs and a final accelerated partition of 100 PFlop/s composed of 364 accelerated nodes based on Nvidia H100 GPUs totaling 1,456 GPUs. Since the autumn, the IDRIS now hosts a small computer called Dalia, which is composed of an NVL72 GB200 from Nvidia.

A NEW ACCESS MODE: STRATEGIC ALLOCATIONS

Strategic Allocations (AS) are dedicated to very large academic and industrial projects capable of strengthening national sovereignty and of strategic interest to France. They are granted on the latest extension of *Jean Zay*.

These Allocations, validated by a strategic committee, are of short duration (a few months maximum), but with very high volumes of hours. The objective is notably to be able to train large-scale AI models in order to meet a very important one-time need while having demonstrated scientific, economic and societal stakes.

SPEAKER DIARIZATION MODEL TRAINING

During the construction phase of the validation process for Strategic Allocation requests, we have selected 6 files from Regular Allocations requesting exceptional resources to propose testing this new type of allocation.
Among the selected projects, the project of the company Pyannote.AI made it possible to train a state-of-the-art diarization model (who speaks when). This model is notably about to be deployed in the video conferencing tool of the new digital suite of the DINUM. This tool will make it possible to offer reliable AI-generated meeting minutes.
Also selected was the project of training new LLMs submitted by Linagora as part of the OpenLLM France research project. This project brings together private and public players, such as Linagora, the CEA, the CNRS or the Class'Code association. 3 LLMs, ranging from 1 billion to 15 billion parameters, have been trained on French and European corpora.

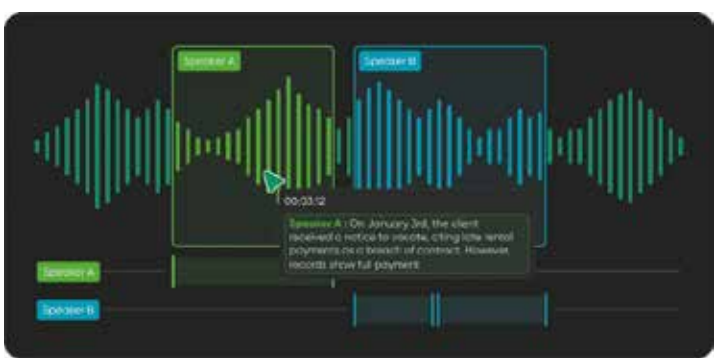


Image of diarization on the *pyannote.AI.* project.

How to make a request?

The submission of a Strategic Allocation request is made on the site www.edari.fr after contacting GENCI to verify the conditions for obtaining it.

Foreword by

Marjorie BERTOLUS

CHAIR OF GENCI'S ASSESSEMENT COMMITTEE, RESEARCH DIRECTOR AND COORDINATOR OF EUROPEAN INITIATIVES AT CEA

In 2025, GENCI's Evaluation Committee supported projects of excellence, illustrating the diversity and dynamism of scientific research.

The year 2025 confirms GENCI's essential role in supporting research requiring high-performance computing resources. In a context of increasing demand, especially for GPU processors, the Evaluation Committee has ensured a balanced and rigorous allocation of resources, serving the scientific quality of the supported projects. As in previous years, the vast majority of requests were met thanks to the joint commitment of the thematic committee chairs, the centre directors and the GENCI teams.

The year is also marked by a growing diversification of uses. Projects relying on GENCI's resources span an increasingly broad spectrum, with a rise in work at the interface between simulation, large-scale data processing, and artificial intelligence, as well as applications in high-impact fields such as health, climate, and engineering. In line with ongoing developments, the deployment of the *Jean Zay 4* machine, the arrival of the quantum computers *Ruby* and *Lucy* and the announcement of the future exascale machine *Alice Recoque* represent major milestones for the French research community. These advances open new perspectives for the various actors and illustrate GENCI's ambition to operate at the highest international level, while supporting users in the transition to these new architectures.

Interactions with users, strengthened in recent years, continue to play a structuring role, through user committees, but also through the "GENCI Users" webinar series, training activities and hackathons.

I would like to thank all those involved for their commitment and the quality of the work accomplished.

Forewords by the chairs of the thematic committees (CT)

Sally CLOSE
Chair of CT1

CT1
ENVIRONMENT

The number of projects reviewed by CT1 during 2025/6 has remained more or less stable compared to recent years, with around 50 projects in total, of which 6 were new this year. The subjects treated by the projects cover the whole of the domain, **even if climate modelling remains a predominant theme, and the results obtained by the users include technical as well as scientific advances.** There is still a strong demand for the CPU, with growing but still fairly limited GPU use, although a number of projets now include tasks that seek to leverage machine learning-based approaches, and some community codes are being ported to GPU. The quality of the dossiers remains very good, with a number of projects receiving funding from the ANR or Europe, and early-career researchers participate in most of the projects.

Rémi ZAMANSKY
Chair of CT2a

CT2A
NON REACTIVE AND MULTIPHASE FLOWS

The equations of fluid mechanics remain impossible to solve analytically, and obtaining accurate numerical solutions to these inherently multiscale problems requires harnessing the full extent of available computing power. High-performance computing has thus become an indispensable tool in fluid mechanics research, enabling the discovery of the fundamental mechanisms of hydrodynamic instabilities, turbulence, and those emerging from "multiphysical" couplings. **The power of supercomputers is also leveraged to generate reference databases essential for the development of new, application-oriented models that are more efficient in terms of computational resources.** The number and diversity of projects (53 proposals in 2025 for over 350 million hours of computing time requested) demonstrate that numerical flow simulation is at the heart of current challenges: energy, the environment, processes…

Stéphane JAY
Chair of CT2b

CT2B
REACTIVE AND MULTIPHASE FLOWS

The projects of CT2b are characterized by the diversity of themes involving multiphase and/or reactive flows, combining highly fundamental work based on direct numerical simulations with more applied configurations close to industrial conditions in the fields of transportation, energy, and the environment. **This is reflected in the proposals submitted in 2025, which aim to enhance energy system security, improve efficiency, and reduce emissions.** The contribution of CT2b to decarbonized energy production over several years has been very significant. A notable highlight is the emergence of lattice Boltzmann approaches, which present challenges in considering multiphysics aspects. Their performance, particularly on GPUs, offers the prospect of substantial gains in simulation capacity in the coming years.

Morgane Dos Santos
Physics researcher

CT3
BIOLOGY AND HEALTH

In 2025, Thematic Committee 3 "Biology and Health" evaluated 12 projects under allocation A18 and 16 projects under allocation A19. **These projects span a range of fields, including medical imaging, radiotherapy, molecular and fluid dynamics, and neuroscience.** Of all the proposals reviewed, 13 were new projects, reflecting the renewal of research themes and teams, while 15 were renewals, illustrating the continuation of ongoing work and the need for sustained access to computing resources. In total, more than 83 million compute hours were allocated. Over 90% of these resources were consumed on the *JOLIOT-CURIE*/Irene Rome and ADASTRA Genoa partitions.
Overall, the committee's activity highlights the critical role of high-performance computing in biology and health, both for advancing the understanding of underlying mechanisms and for developing applications with significant societal impact.

Dubois Yohan
Chair of CT4

CT4
GEOPHYSICS AND ASTROPHYSICS

In 2025, the CT4 saw a sharp rise in demand for computing time: a 50% increase in CPU hours (450 Mh) and an 86% increase in GPU hours (2.3 Mh) across 56 projects. This marks a significant milestone in the adoption of accelerated architectures: many fields have migrated to GPUs (NVIDIA and AMD). However, the transition remains more challenging for certain legacy applications, particularly those based on adaptive mesh refinement. CT4 stands out for the diversity of its approaches (finite volumes, spectral methods, SPH, PIC, MHD, radiative transfer) applied to systems spanning a wide range of scales across twenty orders of magnitude, from the Earth's core to the observable Universe.

RESEARCH COMMUNITY

Virginie Grandgirard
CEA senior research director - CEA fellow

CT5
THEORETICAL PHYSICS AND PLASMA PHYS

In 2025, the number of projects supported by CT5 remained stable with 30 projects. The overall computing volume for our community (magnetic and inertial confinement fusion, cold plasmas, field and particle theory, condensed matter, etc.) grew by 6% compared to 2024, reaching 899 million CPU hours and 8.354 million GPU hours.
CPU partitions remain clearly favored, and CT5 confirms its position as the largest user of the Irene-Rome partition, with 743 Mh, accounting for nearly 33% of the available hours.
At the same time, the shift toward accelerated computing is intensifying, with GPU consumption nearly tripling (a 2.7x factor). **This progress reflects the optimization of QCD and plasma physics codes, which are diversifying their architecture usage by employing not only NVIDIA but also AMD GPUs** (1.89 Mh on MI250X GPUs, representing 19% of the total allocation). While the integration of AI into projects remains marginal, its potential to optimize simulations and analyze massive datasets has yet to be fully demonstrated.

Hélène Barucq
Senior research director

CT6
COMPUTING, ALGORITHMS AND MATHEMATICS

In 2025, 14 projects were submitted, 8 of which were new proposals. It is clear that Technical Committee 6 does not receive a large number of proposals, but some projects bring together multiple research teams. This is an interesting collaborative effort, in line with the goal of creating the future French software stack. **There has been an increase in submissions related to quantum computing, primarily in the field of hardware.** Numerical modeling aided by machine learning and statistical methods is also beginning to gain prominence within this technical committee, with a growing use of GPUs.

Patrick FUCHS
Associate Professor
Paris Cité University

CT7
MOLECULAR MODELLING APPLIED TO BIOLOGY

CT7 evaluated 47 projects in 2025 and allocated approximately 132 million CPU hours (CSL equivalent) and 8 million GPU hours (V100 equivalent). For several years now, all the GPU partitions available at GENCI have been extensively used by CT7, with most of the code being ported to the various architectures.
All-atom, coarse-grained, or QM/MM molecular dynamics remains the core technique for users, applied to all kind of biomolecules (proteins, membrane proteins, nucleic acids, protein/peptide interactions, etc.). Some projects also use tools like AlphaFold2 and 3 to predict protein structure from their sequence. Machine learning is increasingly used to improve sampling or analyze simulations. Again, GENCI's resources are essential to our researchers.

David Loffreda
Chair of CT8
(1st period)

Céline Chizallet
Chair of CT8
(2nd period)

CT8
QUANTUM CHEMISTRY AND MOLECULAR MODELLING

With over 50 applications submitted through regular access and a hundred active projects in dynamic allocation in 2025, CT8 demonstrates growing dynamismin areas related to gas-phase or condensed-phase chemical reactivity and reactive interfaces. The development of codes to meet these objectives with precision and efficiency remains a very active field of research. The year 2025 is marked by the frequent use of tools based on artificial intelligence for the development of interatomic potentials, leveraging GPUs provided by GENCI. Training data is obtained using essential quantum chemistry approaches, relying on CPU resources. The impact of the work of the theoretical chemistry community on fields as diverse as energy transition, corrosion, new material discovery, and the fundamental understanding of chemical, geological, and biological processes could be greatly enhanced.

Michèle RAYNAUD
Senior research director

CT9
PHYSICS, CHEMISTRY AND MATERIALS PROPERTIES

In 2025, the 91 projects submitted to CT9 demonstrate to our community the importance of issues related to materials in terms of the development of methods and studies of various systems. The overall allocation granted to CT is substantial, comprising 531 million CPU hours and 2.6 million GPU hours. The effort to port codes to GPU partitions continues. This is placing increasing pressure and demands on the GPU partitions. These allocated resources have enabled the exploration, using ab initio DFT and molecular dynamics methods, of the spectral, optical, magnetic, thermodynamic, and transport properties of various systems with the aim of identifying, for example, sustainable alternatives to materials as well as to the associated manufacturing processes. **The future is focused on improving the quality of predictions, increasing the scales of simulation, and enhancing the coupling between scales using AI.**

Michaël KRAJECKI
Director of the Research Division of the Ministerial Agency for Defense AI (AMIAD)

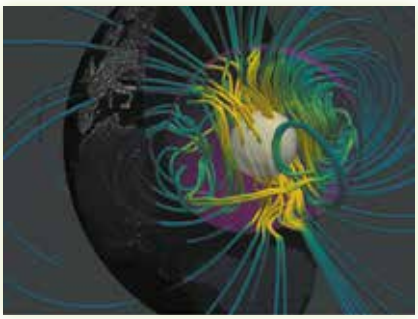
CT10
AI AND CROSS-CUTTING APPLICATIONS OF COMPUTING

In 2025, just like the previous year, the **CT10 saw its number of projects increase by around 30% within the framework of regular access, exceeding 80 evaluated projects. The AI community once again endorses the computing infrastructures provided by GENCI** to develop their research and innovation projects! I would also like to highlight the contribution of strategic access, which was set up in a prefiguration mode and allowed certain projects to accelerate by allocating significant resources on *Jean Zay* over a determined period in "sprint" mode. To mention just one project, I would like to cite the one led by pyannote.AI, which enables the transcription of meetings in the Visio tool led by DINUM. This last word on behalf of CT10 is an opportunity for me to thank you for your trust and to welcome our new experts of CT10 and its new president, Olivier Cappé! Each year, I have had the chance and pleasure of discovering new projects, always more ambitious, and I have no doubt that Olivier will accompany them in the years to come.

RESEARCH RESULTS

THE COMPUTATIONAL CHALLENGE OF MODELING THE GEODYNAMO

Beneath our feet, thousands of kilometers deep, **the Earth's liquid outer core**—an ocean of churning molten iron and nickel—generates **the Earth's magnetic field.** The magnetization of very ancient rocks shows that this magnetic field has existed for about 4 billion years. However, in those ancient times, the current mechanism—convection in the liquid iron core—was likely impossible due to the heat released above by the gradual solidification of the mantle rocks. So where did this magnetic field come from? A recent study (Schaeffer et al. 2025) conducted the first simulations of a dynamo in the magma ocean located at that time between the core and the solid mantle. These simulations required high resolutions and extensive computations on GENCI's GPU-based machines to model this relatively thin fluid layer, and they show that the magma ocean alone cannot explain the ancient magnetic field. Although other possibilities have been proposed, the mystery remains unresolved for now! These simulations and many others can also be used for other studies, even for other



planets, thanks to a brand-new interactive online catalog: the «planetary dynamo simulation explorer» : https://bit.ly/4uvPl1l

Increasing raw computing power will not be enough. We need to improve our models so that we can handle small scales. Although these are not of great scientific interest, they must be included in simulations to obtain a result. Being able to model the effect of these small vortices without calculating them explicitly is crucial for improving the models while reducing their power consumption.

NATHANAEL SCHAEFFER,
SENIOR CODE DEVELOPER *GEODYNAMO XSHELLS*

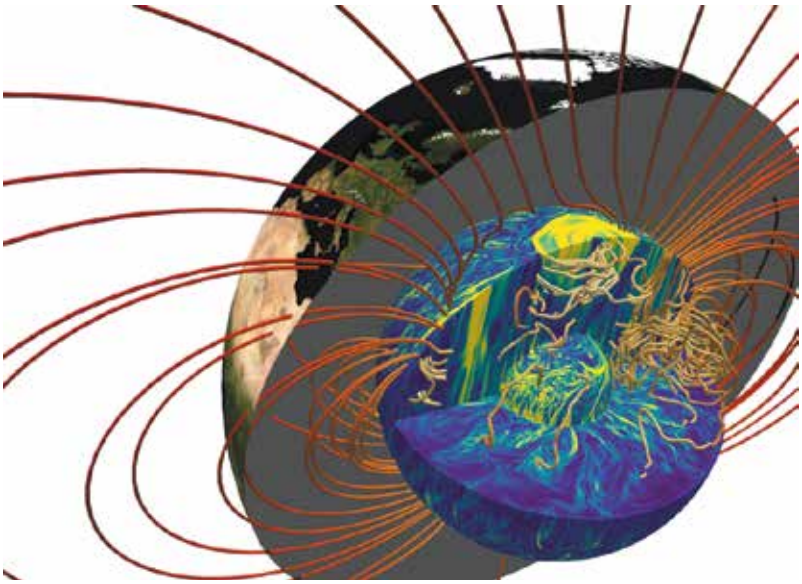
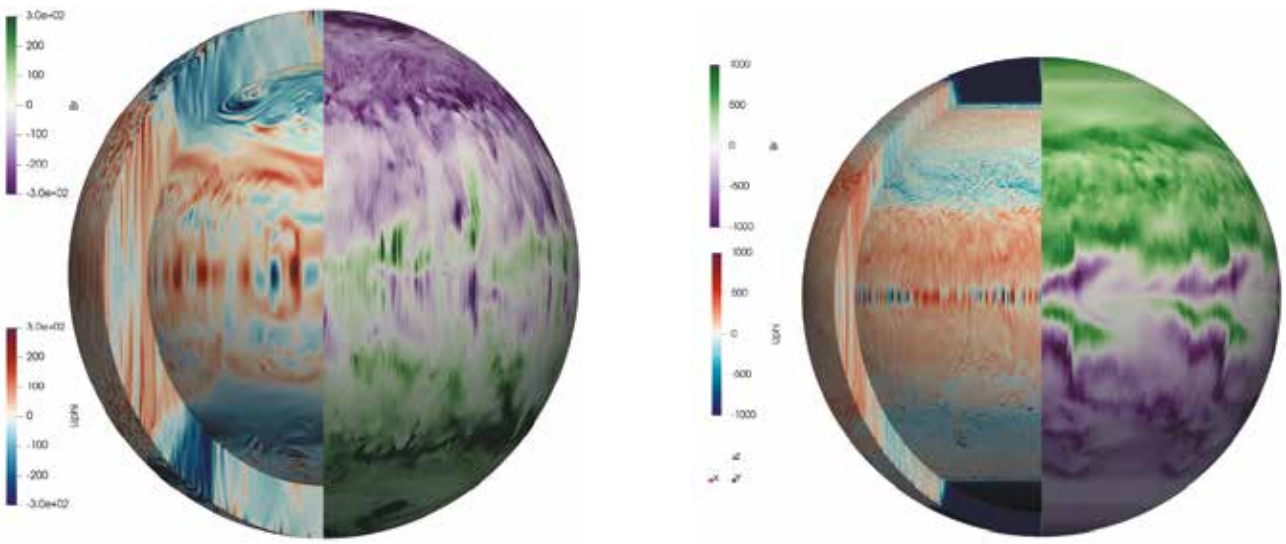


Figure 1: View of the Earth's core from a geodynamo simulation showing **magnetic field lines,** which form a **beautiful dipole at the Earth's surface** but exhibit a complex structure due to their interaction with the movements of the core's molten metal. These movements are represented by colors ranging from dark blue (liquid at rest) to bright yellow (rapid movements). A large vortex is present at the north pole of the core.

Figure 2: 3D snapshot of dynamo simulations in a BMO. The left hemisphere shows **the azimuthal velocity** (turbulent and columnar flows), while the right hemisphere displays **the radial magnetic field** (dipolar structure). Simulations like these require **massive parallel computing power** to resolve fine-scale turbulence. The difference between the two is the thickness of the BMO, which gradually closes as it solidifies from the top.



RESEARCH RESULTS

CT1
Environment

LAERO, Univ. Toulouse, CNRS, IRD

Jean-Pierre CHABOUREAU
Researcher

ADASTRA: 50,000 GPU hours / 20,000 CPU hours *Jean Zay*: 1 million CPU hours

Modeling extreme winds at the heart of tropical storms and tempests. Violent and swirling, extreme winds can cause considerable damage. Our very high-resolution simulations now reveal their dynamics, paving the way for a better understanding and more reliable prediction.

The atmosphere can produce extremely violent and swirling winds capable of causing significant damage. These phenomena, often highly localized, form at very small scales in the atmosphere. They are so small that current weather prediction models cannot represent them accurately.

Simulating them with precision is therefore a major challenge. Indeed, reproducing these gusts and vortices in numerical models requires enormous computing power, which is still difficult to mobilize over the short timescales required for weather forecasting.

To better study them, we used an optimized version of the Meso-NH atmospheric model, designed to run efficiently on the GPU-based ADASTRA supercomputer. This allowed us to perform very high-resolution numerical simulations capable of reproducing tropical storms and convective systems in much greater detail.

This work has led to two major advances. First, gusts and vortices are now better represented in the simulations, providing a more realistic view of their role in tropical storms and convective systems. Second, the computations have become much faster and more energy efficient: the Méso-NH model now runs about six times faster while consuming twice less energy. These improvements open the way to even finer and faster simulations.

Some of these simulations will also be used to build a reference database on tropical convection. This database will support the development of algorithms for a future space mission of the French space agency CNES called C2O-MODO. Ultimately, these advances will improve our understanding of extreme winds and help enhance their prediction, which is essential for the safety of people and infrastructure.

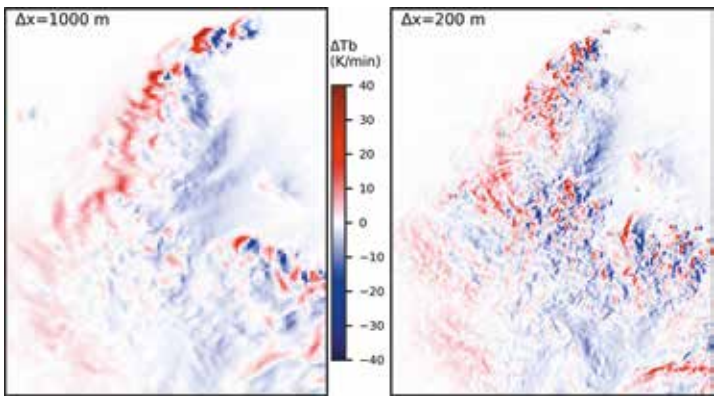


Figure 1: One-minute varation of the 183 GHz brightness temperature over a mesoscale convective system off Cabo Verde. Left, image simulated by Meso-NH at the resolution of current numerical weather prediction models. Right, impact of the fine-scale representation of turbulence. The details highlight the effect of very strong vertical velocities explicitly represented in the convective cores of the system.

CT4
Geophysics and Astrophysics

CEA/AIM

Ugo LEBREUILLY
Researcher

Gabriel VERRIER
Ph. D candidate

Patrick HENNEBELLE
Researcher

Joliot-Curie (Rome): 15 million CPU hours

Interstellar dust grains are the fundamental building blocks of planets. Constraining their collision rate is key to progress in our understanding of planet formation.

The formation of planets is still a great mystery. Indeed, interstellar dust grains, the building blocks of planets, are expected to grow more than 12 orders of magnitude in protoplanetary discs to form planetary cores. Dust coagulation is almost certainly responsible for grain evolution from micron to centimeter while dynamical instabilities allow planet formation from these centimeter-sized grains. Recent observations of forming protostars indicate that coagulation might be significant before the disc is even born, which contradicts the current theoretical predictions of grain growth. To discover a channel that would allow for the rapid formation of these large grains is one of the goals of the simulations of the project "The Dawn of Planets".

Turbulent is most likely to be the main driver of grain growth. "State-of-the-art" theoretical models of grain collisions in a turbulent flow struggle to explain the formation of large grains prior to disc formation. But they neglect a fundamental aspect of grain dynamics: dust can concentrate! In order to improve these models, we have analyzed the dynamical evolution of a gas-dust mixture under forced turbulence with the community-based code RAMSES. Those simulations uncovered a significant dust decoupling under protostar-like conditions. This decoupling could be the key toward faster coagulation as it allows for more frequent grain-grain collisions. With these new simulations, we aim at developing a new theoretical model of grain growth that will be used in the context of planet formation.

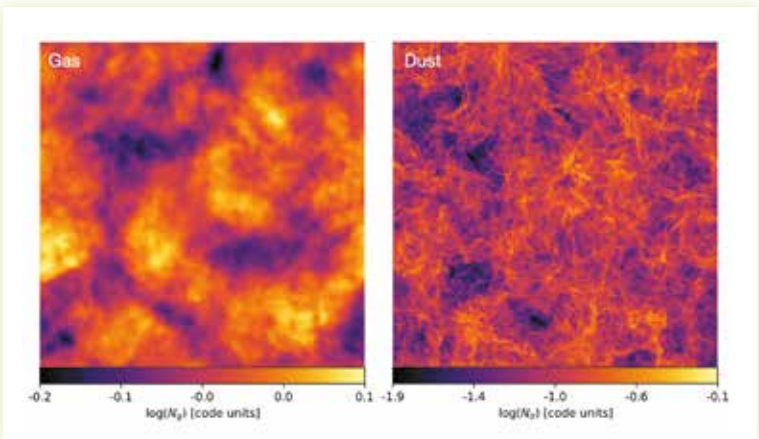


Figure 1: Column density maps of the gas (left) and dust (right) for one of the simulations of the project. We observe a strong decoupling between gas and dust. Dust concentrates favorably in certain regions of the flow, which favor grain-grain collisions.

RESEARCH RESULTS

CT2A
Non reactive and multiphase flows

Laboratoire de Mécanique des Fluides de Lille – Kampé de Fériet

Meldi MARCELLO
Professor Arts et Métiers, Lille

Joliot-Curie: 2 million CPU hours

This project improves airflow prediction in axial compressors by combining simulations, experimental measurements, data assimilation, and machine learning. A multi-fidelity approach is conceived and performed thanks to the usage of high performance computing facilities.

This project targets how to better predict the complex airflow inside an axial compressor, a machine used in engines for aerospace applications. Although computer simulations have become very powerful, they still struggle to capture all the details of turbulent flow in real industrial conditions. To improve accuracy, we propose to combine simulations with real measurements taken in a laboratory using laser-based flow imaging. This coupling is performed via Data Assimilation, which blends numerical model predictions with experimental data so the simulation can be corrected as it runs. The specific technique, the Ensemble Kalman Filter, normally requires many high-precision simulations, which are very expensive. To reduce the cost, we use a multi-fidelity approach that mixes fast, coarse simulations with computationally expensive, more detailed ones. This complete dataset accounting for numerical prediction and experimental results is used to train neural networks models, which complement the numerical system increasing accuracy and reducing further computational costs. The complete strategy is shown in Figure 1. Calculations were performed using the computational resources of the IRENE cluster, for a total of 2 million scalar hours. The simulations within the ensemble were executed using different numbers of cores, depending on the size of each numerical problem. Specifically, the low-fidelity simulations used grids of 5 million elements and were run on 96 cores, each of them requiring 2,300 scalar CPU hours. the high-fidelity simulations employed grids of 50 million elements and required 640 cores (400,000 scalar CPU hours).

Early results suggest that this hybrid strategy based on physics-based modeling combined with machine learning could offer a more efficient way to predict airflow in complex machines, helping engineers design better and more reliable compressors.

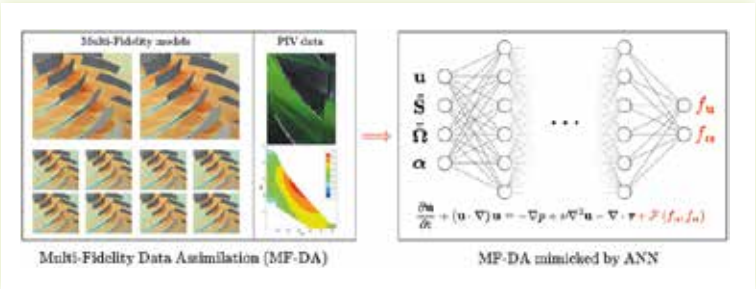


Figure 1: Scheme of the multi-fidelity data assimilation methodology (left) and the approximation of this process through the integration of an artificial neural network to predict a forcing term in the dynamic equations (right).

CT 2B
Reactive and multiphase flows

IFP Energies nouvelles

Karine TRUFFIN
Research Engineer

Joliot-Curie (SKL and Rome partitions): 7,861,835 CPU hours

GENCI computing resources enabled high-fidelity numerical simulations to investigate hydrogen combustion and advanced energy systems. The work combines *LES* and DNS simulations with uncertainty quantification to improve the understanding of physical mechanisms in hydrogen engines and other energy applications.

This project aims to develop numerical tools and models to improve the understanding and prediction of complex mechanisms occurring in hydrogen engines and advanced energy systems. Simulations were mainly performed using Large-Eddy Simulation (*LES*) and Direct Numerical Simulation (DNS) with the CONVERGE CFD solver.

A first part of the work focused on hydrogen combustion in ultra-lean spark-ignition engines. Multi-cycle *LES* were conducted on an optical single-cylinder engine to analyse interactions between turbulence, ignition and flame propagation. The simulations reproduce the main flow structures observed experimentally as well as the evolution of the in-cylinder pressure and its cyclic variability. They nevertheless highlight the need for predictive combustion models applicable across different turbulence levels and fuels.

An original methodology combining *LES*, experimental diagnostics and uncertainty quantification was also developed. Multi-fidelity surrogate models were used to evaluate the influence of parameters such as spark timing and combustion model parameters while limiting the computational cost of *LES* campaigns.

GENCI resources were also used to generate a high-fidelity database dedicated to machine learning. DNS of a turbulent premixed hydrogen slot burner produced detailed three-dimensional thermo-chemical fields. These datasets are used to train convolutional neural networks aimed at improving the modelling of subgrid-scale turbulent combustion terms in *LES*.

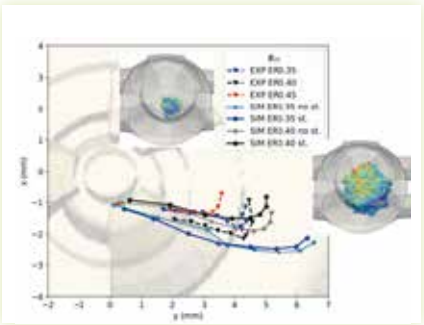


Figure 1: Flame kernel Centre trajectory. Comparisons between experimental results of a spark-ignition engine (triangles) and *LES* (circles) for 2 equivalence ratios (0.35 and 0.4).

In addition, new numerical approaches were developed. An Adaptive Mesh Refinement (AMR) strategy was proposed to optimise the resolution of turbulent structures in hydrogen flame acceleration simulations while reducing computational cost.

Finally, CFD simulations were performed to investigate the hydrodynamic behaviour of floating offshore wind turbine platforms within the international OC7 benchmark framework. Three-dimensional simulations reproduce platform motions and hydrodynamic loads, demonstrating the relevance of high-fidelity CFD for the design of next-generation offshore wind systems.

RESEARCH RESULTS

CT5
Theoretical physics and plasma physics

Joliot-Curie (Rome): 2,000,000 hours

Particle-in-Cell (PIC) simulations, which are commonly used in plasma physics, require considerable computational resources. The Sparse-PIC project introduces a new class of PIC algorithms based on a combination of sparse grid techniques and PIC algorithms to reduce computational resources.

Laboratoire Plasma et Conversion d'Energie, LAPLACE, Univ. Toulouse, CNRS, INPT

Laurent GARRIGUES
Senior scientist at CNRS

Maison de la Simulation, CEA, CNRS, Univ. Paris Saclay

Marc CHUNG-TO-SANG
Postdoctoral researcher

Mathieu LOBET
CEA
Research-Engineer

INRIA Bordeaux

Clément GUILLET
Researcher

Inst. Mathématiques, IMT, Univ. Toulouse, CNRS
Fabrice DELUZET
Research-Engineer at CNRS
Jacek NARSKI
Associate Professor, P. Sabatier Univ.

The simulation under real conditions of partially magnetized low temperature plasmas by Lagrangian approaches, though using powerful Particle-In-Cell (PIC) techniques supplemented with efficient high-performance computing methods, requires considerable computing resources for large plasma densities. This is explained by two main limitations. First, stability conditions that constrain the numerical parameters to resolve the small space and time scales. Second, PIC methods rely on a sampling of the distribution function by numerical particles whose motion is time integrated in the self-consistent electric field. The PIC algorithm remains close to physics and offers an incomparable efficiency with regard to Eulerian methods, discretizing the distribution function onto a mesh. It is widely and successfully operated for the discretization of kinetic plasma models for more than 40 years. Nonetheless, to spare the computational resources, the number of numerical particles is limited compared to that of the physical particles. Inherent to this "coarse" sampling, PIC algorithms produce numerical approximations prone to statistical fluctuations that vanish slowly with the mean number of particles per cell. A breakthrough is therefore necessary to reduce the computational resources by orders of magnitude and make possible the use of explicit PIC method for large scale and/or densities for 3D computations.

The ANR MATURATION project, of which this study is a part, aims to introduce a new class of PIC algorithms with unprecedented computational efficiency, by analyzing and improving, parallelizing and optimizing as well as benchmarking, in the demanding context of low temperature plasmas through 2D large scale and 3D computations, a method recently proposed in the literature, based on a combination of sparse grid techniques and PIC algorithm.

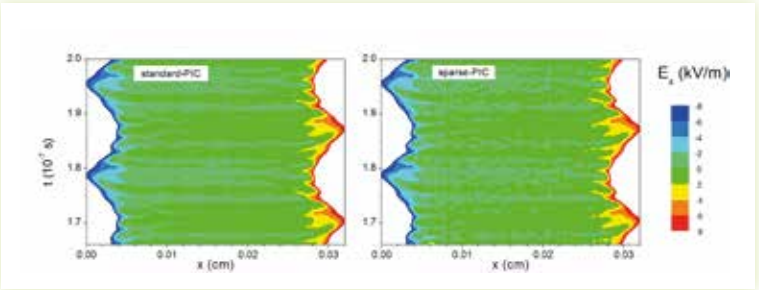


Figure 1: Comparison of the axial electric field profile as a function of time obtained with the Standard and Sparse PIC algorithms in 2D simulations of an RF capacitively coupled plasma discharge. The mesh size is 5122, with 50 million and 5 million particles for the Standard and Sparse PIC methods, respectively. The Sparse-PIC method is capable of reproducing the sheath instabilities observed with a computational speed up factor of 4. This work was funded by l'Agence Nationale de la Recherche (ANR) under the Project No. ANR-22-CE46-0012.

CT9
Physics, chemistry and materials properties

Joliot-Curie (Rome): 700,000 hours

As part of a collaboration between the CNRS and CEA in Grenoble, we have reduced the cost associated with a family of approaches, labeled many-body perturbation theories in quantum mechanics, by several orders of magnitude for the study of the optoelectronic properties of crystalline or organic systems comprising up to several thousand atoms.

Institut Néel, CNRS/UGA, Grenoble

Xavier BLASE
CNRS Research Director

Laboratoire L_SIM/MEM/IRIG/CEA Grenoble

Ivan DUCHEMIN
Research Engineer CEA

The description of the electronic properties of systems of interest in physics and chemistry relies on quantum mechanics, whose complexity grows exponentially with the number of electrons. The development of associated numerical tools combines two complementary strategies: the writing of modern, massively parallel codes (e.g., MPI/OpenMP), and jointly, the development of approximations and algorithms that reduce the number of operations required to calculate a physical observable.

As part of a collaboration between CNRS and CEA Grenoble, we have developed the codes Fiesta (French Initiative for the Electronic Structure of Thousands of Atoms) and BeDeft (Beyond Density Functional Theory), which implement many-body perturbation theories offering a very good compromise between accuracy and computation time. The reformulation of quantum mechanics operators (e.g., electronic response functions) in a so-called real-space and imaginary-time approach reduces the cost of these approaches, with the number of operations required now growing as the cube—rather than the fourth power—of the number of electrons in the system. On the other hand, the development of "Russian dolls" approaches allows treating an "active" subsystem of interest with high precision, while taking into account the effect of a complex environment (a solvent, a metal or semiconductor substrate, etc.), whose complexity can be reduced by partitioning and automatic compression to a few effective degrees of freedom. Our applications have focused on disordered organic systems for photovoltaics or flat screens (OLEDs), but also crystalline systems with defects (quantum emitters in hexagonal boron nitride). Our work strives to find a balance between the growth of computing resources and the need to reduce consumption.

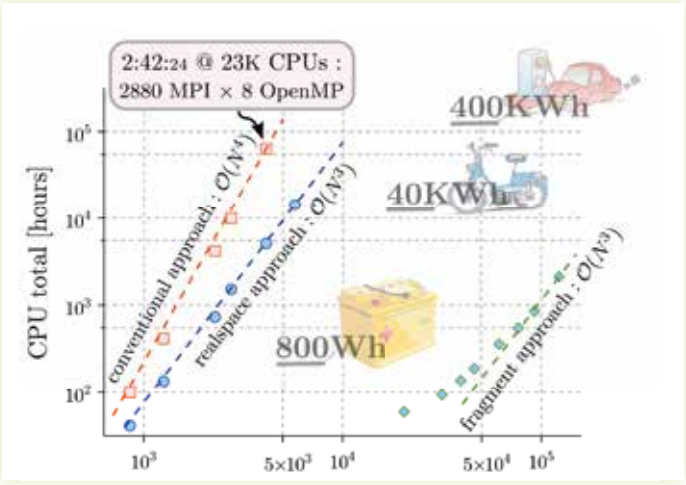


Figure 1: CPU time as a function of the number of electrons for calculating electronic properties using the GW approach (G Green's function, W screened Coulomb potential). The traditional approach with quartic complexity (red) is compared to an approach with cubic complexity (blue) and a fragment approach (green) with cubic complexity but with a reduced prefactor through decoupling into subsystems. The approximate conversion of computing time into power consumption is schematized (Image Ivan Duchemin, L_SIM, IRIG, CEA/Grenoble).

RESEARCH RESULTS

CT8
Quantum chemistry and molecular modelling

Laboratoire CPCV, Département de Chimie ENS-PSL, SU, CNRS

Guillaume STIRNEMANN
Research Director and Adjunct Professor

Zakarya BENAYAD
PhD Student

Olaia ANTON
PhD Student

Pierre VIEILLARD
PhD Student

Laboratoire CPCV, Département de Chimie ENS-PSL, SU, CNRS & LMCE, CEA, Univ. Paris Saclay

Mathilde KRETZ
PhD Student

Jean Zay, *Joliot-Curie* and ADASTRA: 2 millions CPU and 165,000 GPU hours

Biological function relies on molecular motion, yet most approaches remain static. Using *Jean Zay*, we built DynaRepo, a large-scale molecular dynamics database, and developed DynamicGT, a dynamic-aware deep learning architecture that improves the prediction of biomolecular interactions, especially in flexible and disordered regions.

The emergence of artificial intelligence is profoundly transforming computational chemistry, particularly through machine-learning interatomic potentials (MLIPs). Unlike traditional representations of potential energy surfaces, MLIPs overcome conventional scaling limitations by providing an effective balance between accuracy and computational efficiency in calculating atomic energies and forces for molecular simulations. As a result, they have significantly advanced simulations across a wide range of applications, including large-scale modeling of materials, interfaces, and chemical reactions.

Despite these advances, the construction of high-quality training datasets, essential for MLIP accuracy, remains a major challenge, especially in the context of chemical reactivity. Reactive processes often involve rare barrier-crossing events that are difficult to capture and therefore underrepresented in standard datasets. To address this issue, our group has co-developed ArcaNN, a comprehensive framework for generating training datasets tailored to reactive MLIPs. ArcaNN combines active learning with advanced sampling strategies to ensure accurate representation of high-energy configurations. It integrates automated workflows for iterative training, exploration, configuration selection, and energy and force labeling.

We have applied this framework to investigate molecular mechanisms, as well as the thermodynamic and kinetic properties, of several key chemical reactions in aqueous environments, particularly those relevant to the origins of life and the potential roles of catalysts and activating groups. Ongoing work extends this approach to a variety of systems, including homogeneous catalysis for CO_2 reduction and energetic materials. In parallel, we are exploring new methodological directions that integrate generative AI models with active learning to further accelerate and optimize the construction of MLIP training datasets.

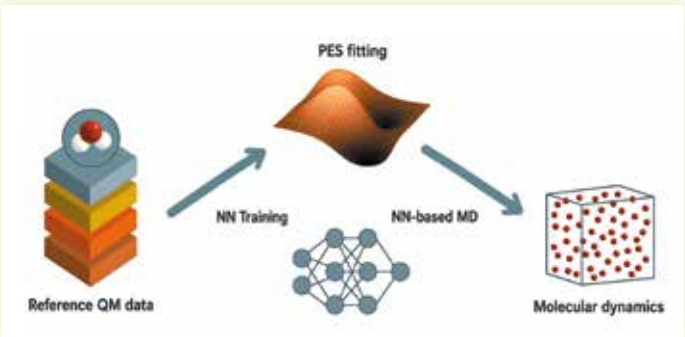


Figure : Machine-learning interatomic potentials (MLIPs) consist of training a machine learning model, typically a neural network (NN), to reproduce reference quantum data by mapping local atomic environments to the system's potential energy surface. These models can then be used to efficiently propagate molecular dynamics simulations of the system.

CT7
Molecular modelling applied to biology

Jean Zay: 404,000 GPU hours H100 and 180,000 GPU hours V100

Biological function relies on molecular motion, yet most approaches remain static. Using *Jean Zay*, we built DynaRepo, a large-scale molecular dynamics database, and developed DynamicGT, a dynamic-aware deep learning architecture that improves the prediction of biomolecular interactions, especially in flexible and disordered regions.

Inria

Yasaman KARAMI
Research scientist

Hamed KHAKZAD
Junior Professorship Chair

Omid MOKHTARI
PhD Student

Biomolecules such as proteins, RNA, and DNA are inherently dynamic: their motions govern their function, interactions, and involvement in many diseases. However, most current approaches, including artificial intelligence methods, still rely on static representations, limiting their ability to capture these essential mechanisms.

Using the *Jean Zay* supercomputer, we took a major step forward by generating and exploiting large-scale dynamic data. We developed DynaRepo, the first official MDDB (Molecular Dynamics Data Bank) node in France and now the second largest in Europe, available online (https://inria.mddbr.eu/). This database contains over 700 biological systems and more than 1.1 milliseconds of molecular dynamics simulations. The data are standardized, enriched with precomputed analyses, and shared following FAIR principles, enabling broad reuse by the scientific community.

Building on this resource, we developed DynamicGT, the first deep learning architecture explicitly designed to incorporate molecular dynamics into binding site prediction. Unlike traditional approaches, DynamicGT leverages conformational ensembles derived from simulations, allowing it to better capture flexible and disordered regions that are often missed by static structures. This approach significantly improves performance while reducing the need for large training datasets. The code is openly available (https://github.com/amisteromid/DynamicGT).

These results demonstrate that integrating molecular dynamics is a key step toward the next generation of AI models in biology. They open new perspectives in drug discovery, protein engineering, and the detailed understanding of molecular mechanisms, bridging structure, dynamics, and function.

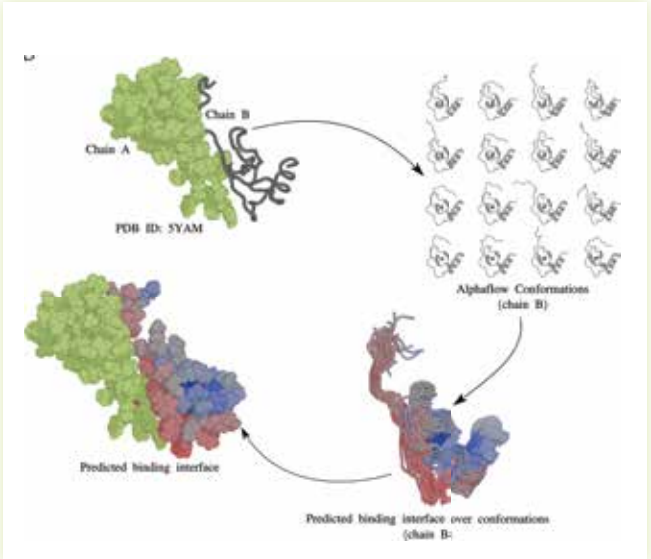


Figure 1: Example of molecular dynamics analysis used in DynaRepo and exploited by DynamicGT: conformational ensembles capture biomolecular flexibility and reveal interaction sites that remain hidden in static structures. (Figure adapted from DOI: 10.1016/j.cels.2025.101454).

RESEARCH RESULTS

CT3
Biology and health

Jean Zay: 28,000 GPU hours A100 65,000 GPU hours H100

Gleamer

Gonzague DE CARPENTIER
AI engineer

Julien GENZLING
AI engineer

Gabriel MISRACHI
VP AI

Radiological images have characteristics that differ fundamentally from natural photographs, creating a domain gap that limits the performance of general-purpose AI models. We develop foundation models trained specifically on large radiological corpora, and apply them to pathology detection and report generation.

Medical images present a particular challenge for AI: their distribution differs profoundly from natural images due to acquisition physics, anatomical structures, and subtle clinical abnormalities. We developed three radiology-specific foundation models—ChestDINO for chest X-rays, BreastDINO for mammography, BoneDINO for musculoskeletal radiography—self-supervised via DINOv2 on millions of clinical radiographs from European and American clinical sources.

Jean Zay's compute allocation enabled a systematic study comparing these models to generic baselines at two scales—ViT-Base and ViT-Large—across frozen probing and end-to-end detection fine-tuning. Our central finding is that domain-specific pre-training is more beneficial at larger scale, with consistent gains across all three modalities. The advantage amplifies as labeled data decreases, making the combination of large architecture and domain-specific pre-training especially valuable for rare pathologies or new clinical sites. Results are confirmed on seven public benchmarks spanning all three radiology modalities. In a second strand, we leveraged these encoders to build ChestVLM, a vision-language model for automatic radiology report generation. Fine-tuned on a large multilingual corpus, ChestVLM outperforms open-source state-of-the-art models on all evaluation metrics.

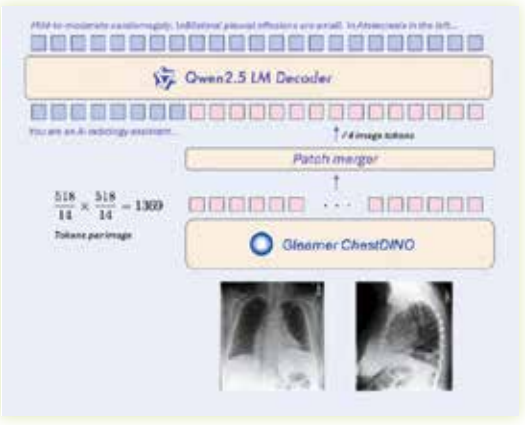


Figure 1: ChestVLM pipeline. Bottom: two chest X-rays (frontal and lateral views) are processed by the ChestDINO encoder (lower central block); the formula 518/14 × 518/14 = 1,369 shows the number of tokens produced per image. The patch merger (central block) reduces this by a factor of four before decoding by Qwen2.5 (upper block). A sample generated report is shown at the top of the figure.

CT6
Computing, algorithms and mathematics

Jean Zay: 40,000 hours 100 50,000 hours CSL

This project explores whether artificial intelligence can develop mathematical intuition and help researchers tackle difficult theoretical problems. Early results show that it can already propose new solutions to some of them.

CERMICS, CNRS, ENPC, Inst. Polytechnique de Paris

Amaury HAYAT
Researcher / Full Professor

François CHARTON
Professor

Axiom Math

Alberto ALFARANO
Professor

This project explores an ambitious question: can artificial intelligence develop genuine mathematical intuition and help solve problems that humans do not yet know how to tackle? The idea is to train deep learning models, in particular *Transformers*, on carefully constructed mathematical datasets so that they learn to propose solutions to difficult theoretical problems. The goal is not to replace mathematicians, but to provide them with a new tool: an AI capable of formulating strong conjectures that can then be rigorously verified. This approach is especially promising in situations where finding a solution is very difficult, while checking whether a proposed solution is correct is much more accessible.

The first results are very encouraging. In particular, the project has focused on the search for *Lyapunov* functions, mathematical objects that are essential for studying the stability of dynamical systems. Intuitively, a *Lyapunov* function plays the role of a kind of "energy": if one can construct a quantity that remains positive and decreases over time, then one can show that the system returns to a stable state rather than diverging. Such functions are fundamental in mathematics, physics, and control theory, but they are often extremely difficult to discover. By training a model on carefully generated examples, we show that it can propose very strong candidate functions. The trained models achieve excellent performance, outperform existing algorithmic solvers and, on some benchmarks, human performance. They even succeed in proposing new candidate functions in cases where no general method is known. This project thus aims to open up a new path at the interface between artificial intelligence and mathematical research, accelerating discovery and bringing forward new mathematical ideas.

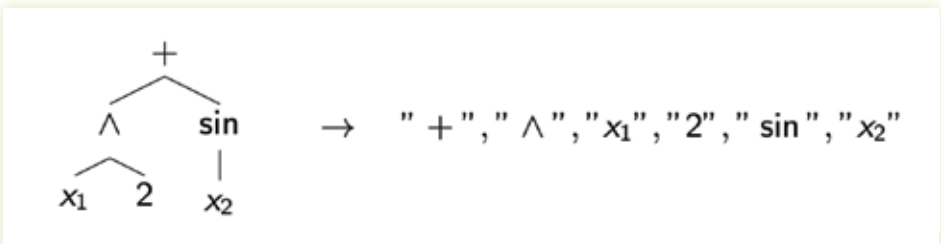


Figure 1: Representation of a mathematical function as the input and output of the neural network (the function is represented as a tree that is enumerated in topological order). This is the format that the neural network "sees" and returns.

RESEARCH RESULTS

CT10
Artificial Intelligence and Cross-Cutting Applications of Computing

Jean Zay: 10,000 hours V100 50,000 hours A100 50,000 hours H100

Dragon LLM

Gaëtan CAILLAUT
Researcher

Raheel QADER
Researcher

Jingshu LIU
Researcher

Mariam NAKHLÉ
PhD student

The objective of this study is to explore methods for adapting a Large Language Model (LLM) to specific business domain cases, particularly the financial domain and the tasks generally associated with it.

We fine-tuned five models with parameter counts ranging from 8 to 70 billion. These training sessions were conducted on a dataset of over 2 million documents and conversations, distributed across various thematic categories with a strong focus on the financial sector. This diversity allows for effective adaptation of the models to the finance domain while preserving their initial capabilities, such as their world knowledge, reasoning abilities, and agentic functionalities. The data breakdown is as follows:

- Finance: 54%
- Translation: 20%
- General: 16%
- RAG (Retrieval-Augmented Generation): 8%
- Reasoning, Mathematics, and Code: 2%

The models are published on the Huggingface platform under licenses that freely permit their use for research purposes. A comprehensive scientific and technical report is also available on arXiv: "The LLM Pro Finance Suite: Multilingual Large Language Models for Financial Applications"

1. Two models with 8 billion parameters (Llama Open Finance 8B and Qwen Open Finance R 8B)
2. A model with 12 billion parameters (Gemma Pro Finance 12B)
3. A model with 32 billion parameters (Qwen Pro Finance R 32B)
4. A model with 70 billion parameters (Llama Pro Finance 70B)

Our models consistently outperform their competitors (of similar size) on financial and financial translation tasks, highlighting the effectiveness of our training strategy, which combines synthetic data generation, domain-specific data collection, and rigorous data cleansing. Furthermore, this specialization does not compromise general-purpose capabilities: LLM Pro Finance models maintain performance similar to leading general-purpose LLMs on standard benchmarks.

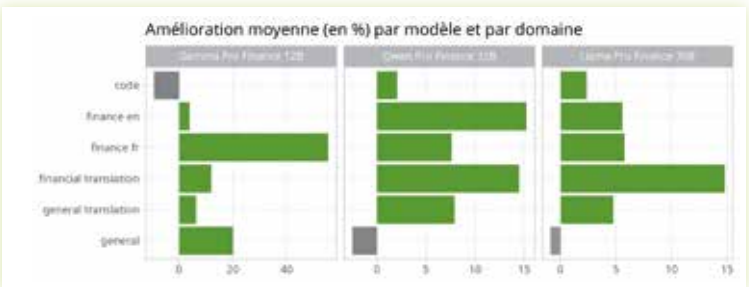


Figure 1: Our specialized finance models perform up to 50% better on financial tasks than the models they were finetuned from, without any significant degradation in performance on more general tasks. The improvements are more pronounced in smaller models, suggesting that it is best to specialize smaller models for use in specific business tasks.

COMMUNITY OF LARGE INDUSTRIAL GROUPS

High fidelity *LES* simulation of an *Openfan* turbine

The RISE (Revolutionary Innovation for Sustainable Engines) technology demonstration program, launched in June 2021, aims at developing new technologies for the next single-aisle aircraft engine and targets a 20% reduction in fuel consumption and CO_2 emissions compared to current commercial engines.

The so-called "*Openfan*" architecture enables a significant reduction in fuel consumption by increasing propulsive efficiency, especially the fan and stator efficiency. Besides their geometries, one of the levers for improving aerodynamic performance is to achieve a portion of laminar flow on the blade.

BETWWEEN LAMINAR FLOW AND TURBULENT

Laminar flow is generally desirable because pressure losses and wall friction are lower, resulting in better performance. In the case of airflow around a propeller, these phenomena take place within the boundary layer (close to the wall), i.e., at a relatively small scale. In this flow, there are usually both laminar and turbulent regions.

The calculations carried out as part of these activities with the Fast code, developed by Onera, aim to perform a high-fidelity simulation of an Openfan with the LES (Large Eddy Simulation) method, to quantify the credit of laminarity, to precisely identify laminar and turbulent areas and the type of transition, and to compare this with practices used in design. Despite its reduced scale, computing all 26 blades of the system using an LES approach remains very costly. The blade modeled with LES contains 90% of the 3 billion computational cells.

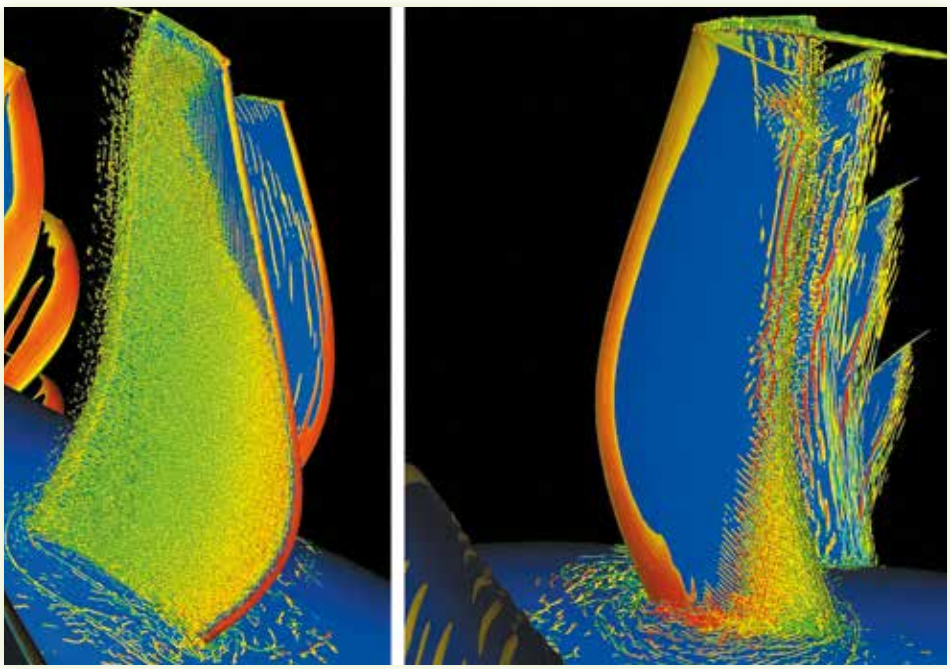


Figure 1: Q Criterion on the suction side (right) and the pressure side (left) showing laminar and turbulent regions.

RESEARCH RESULTS

From academic research to voice intelligence: *pyannoteAI* journey

Born in academic laboratories and powered by GENCI's high-performance computing infrastructure, pyannoteAI demonstrates how rigorous fundamental research can be transformed into a sovereign industrial product. Now integrated into La Suite de la DINUM, the French startup's speaker diarization technology was built on years of model training conducted on the *Jean Zay* supercomputer.

There is something genuinely unusual about pyannoteAI's story: the product was not designed for the market. It was built to answer a precise scientific question — *who speaks, and when?* — in real-world audio recordings that are noisy, overlapping, and far from ideal. It is this commitment to scientific rigour, inherited from ten years of academic research including at CNRS, that constitutes the startup's core competitive advantage today.

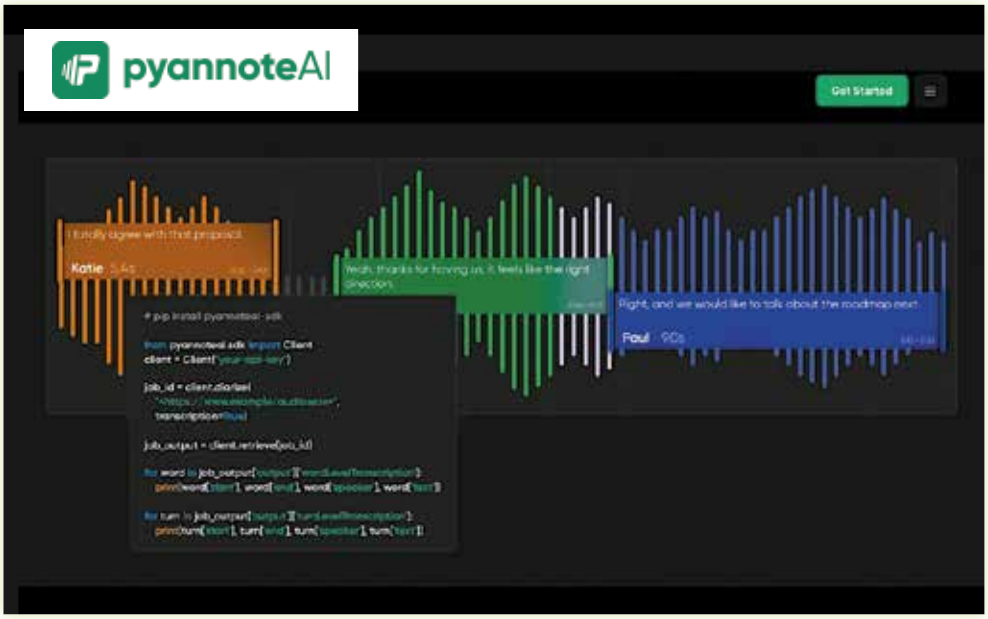
A TECHNOLOGY FORGED IN THE LAB

Speaker diarization — the automatic segmentation of an audio recording according to who is speaking — is a long-standing problem in the speech processing community. What sets pyannoteAI's approach apart is a refusal to cut corners. For years, models were trained, evaluated, and refined within an open research framework, producing peer-reviewed publications, public benchmarks, and a global academic user community that stress-tested the algorithms in the most demanding real-world conditions. The open-source pyannote.audio library, which underpins the entire product stack, has now accumulated more than a billion of downloads. It is deployed across dozens of countries, in applications ranging from medical transcription and forensic analysis to enterprise conversation intelligence and automatic subtitling. This is not coincidence — it is the outcome of a decade of rigorous, iterative improvement.

Hervé Bredin (left) and Vincent Molina (right), the co-founders of the startup *pyannoteAI*.

THE DECISIVE ROLE OF HIGH-PERFORMANCE COMPUTING

The transition from research prototype to industrial-grade product would not have been possible without access to computing resources commensurate with the ambition. Training the latest generation of models — Precision-2 and Community-1 — required thousands of GPU hours on the *Jean Zay* supercomputer at IDRIS, under allocations granted by GENCI. Those resources did not simply enable the training of larger models. They made possible a systematic exploration of architectures, training datasets, and optimization strategies that would otherwise have been out of reach. The ability to iterate rapidly at scale is precisely what allowed pyannoteAI to reach the accuracy levels now recognized as best-in-class on international reference benchmarks. *"Without access to Jean Zay, we could not have run the training experiments that made the difference. High-performance computing is not an accelerator for us — it is a prerequisite."*



Example of a diarization model developed by the startup *pyannoteAI*.

Without access to *Jean Zay*, we could not have run the training experiments that made the difference. High-performance computing is not an accelerator for us — it is a prerequisite.

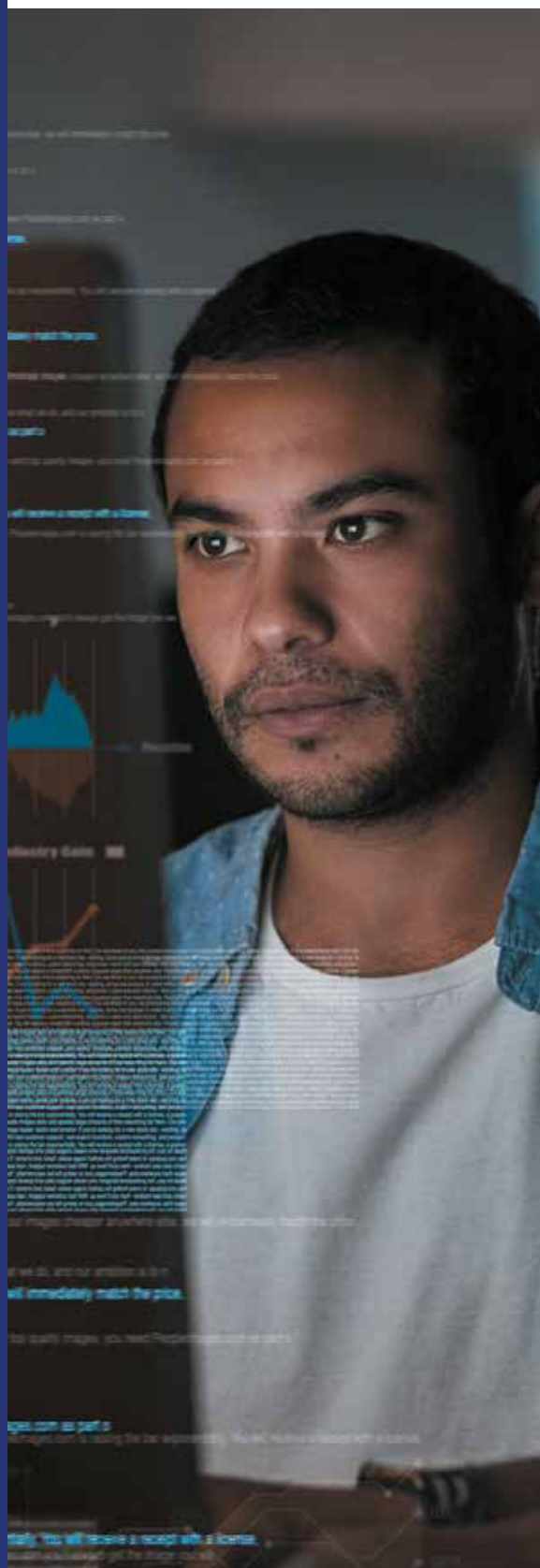
Hervé BREDIN, CHIEF SCIENCE OFFICER AND CO-FOUNDER OF *PYANNOTEAI*

FROM DEEPTECH STARTUP TO SOVEREIGN TOOL

PyannoteAI is today a commercial company, but it has not abandoned its academic roots. The team continues to publish, contributes to public benchmarks, and maintains the open-source dimensions of its ecosystem. This hybrid positioning — bridging scientific rigour and industrial requirements — is what led the French government's digital directorate (DINUM) to integrate pyannoteAI's models into La Suite, the sovereign digital platform for the French public administration.
This integration carries both symbolic and strategic weight: models trained on France's public high-performance computing infrastructure, developed in a spirit of open research, now power the everyday digital tools of the French civil service. The circle is complete — and it is entirely sovereign.lle est entièrement souveraine.

This work was granted access to the HPC resources of IDRIS under allocations 2024-AD011014274R2, 2024-AS011016439 and 2024-A0181016176, made by GENCI.

GENCI AND ITS ENVIRONMENT



CONTENTS

Partner of the MesoNET project in the region

The MesoNET project (2021-2027) consists of a federation of regional mesocenters. Its objective is to strengthen the structuring of national and regional offerings in numerical simulation, through high-performance computing (HPC), combined with artificial intelligence (AI) and quantum computing. Its distinctive feature is local support.

The ministry of Higher Education and Research is seeking to establish a Regional Research Infrastructure (RI) to structure and sustainably provide a digital service offering for computing and processing scientific data for French higher education and research institutions (ESR), integrated into an international framework. This solution will have the advantage of, firstly, streamlining public investment in human and financial resources for the digital sector within ESR and, secondly, offering national and European users, via EGI and EOSC, a single, easily accessible regional service catalog to which local initiatives can contribute.

IMPLEMENTATION OF A COMPLETE OFFER

A proposal was submitted in 2025 by the partners of MesoNet and France Grilles to pool complementary services for the benefit of the scientific and teaching communities. For the past three years, these stakeholders have been building a comprehensive offering through a federated operational structure (shared, streamlined, and based on a principle of devolution) that covers users' needs for massive data processing of HPC, HTC, AI, and quantum computing, from access to regional computing and storage resources and support to access to national and European resources. It also aims to address the shortage of national digital skills by providing services necessary for initial and continuing training, maintaining a network of experts distributed across the country, and fostering interactions, skills, and complementarities.

All the actors of the MesoNET project.

Focus

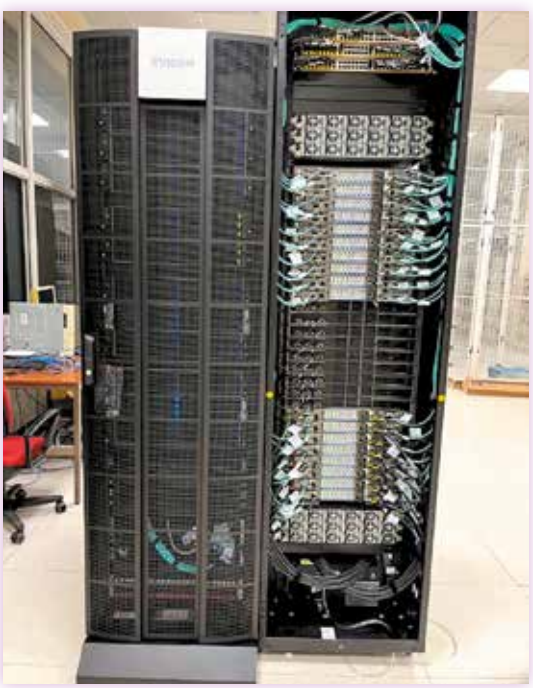
key figures

For the entire Mesonet project, approximately **160 technical personnel** are involved in various working groups, including security, storage, the access portal, and support, with its shared team of engineers. More than **720 users from all regions of France** have already benefited from this equipment in over 160 research projects.

CLUSSTER, sovereign Cloud for private and public partners

As part of the national Cloud strategy, the *CLUSSTER* project aims to establish a sovereign cloud of infrastructure and services for AI between public and private actors.

The *CLUSSTER* project (Cloud Unifié Souverain de Services, de Technologies et d'infrastructuRes), launched in early 2023, was completed on December 31, 2025. It is part of the CLOUD acceleration strategy of the Direction Générale des Entreprises (PIA4) and is funded by BPI France.
Through a public/private partnership, *CLUSSTER* has demonstrated its ability to implement a solution that brings together sovereign, secure, unified, and scalable cloud services (Infrastructure as a Service) in response to needs in Artificial Intelligence (AI), HPC, and quantum computing. It provides a single point of entry for the entire academic and industrial community to a set of infrastructures and services.
In 2025, the DALIA computer, acquired by GENCI, was installed at the CNRS-IDRIS national computing center in an area outside the ZRR (Zone à Régime Restrictif) and put into production after its proper functioning was validated by a major challenge project.
Connections between the eDARI site, which allows applications for open research to be submitted, and the CLUSSTER portal were developed and qualified through functional testing.
The *CLUSSTER* portal will not be put into production in 2026 for either confidential or open research. None of the project's private partners wished to commit to marketing the portal.



The DALIA supercomputer at IDRIS.

Focus

DALIA supercomputer

DALIA consists of 72 NVIDIA GPUs called *Blackwell*, which offer 2.5 times the performance for model training tasks and almost 30 times the performance for inference tasks thanks to the introduction of new 4-bit floating point (FP4) operators. Resource requests are made via eDARI.

EPICURE, support to communities

EPICURE (*Eurohpc aPplICation sUpport pRojEct*) is a project funded by EuroHPC. Its mission is to set up a Europe-wide distributed Application Support Team for high-performance computing.

The EPICURE project entered its second year in 2025. Funded by EuroHPC, it brings together researchers and experts from all the computing centers hosting EuroHPC supercomputers.
Since its kickoff on February 1, 2024, EPICURE experts have provided support to more than 90 projects for code porting or optimization.
France is represented in this project by GENCI and CINES, as an affiliated entity of GENCI.

All projects allocated resources on a EuroHPC supercomputer are eligible for EPICURE support.

Annual Meeting 2025.

Eupex is evolving but continues on its path

The EUPEX (*EUropean Pilot for EXascale*) project, launched in 2022 and co-funded by the EuroHPC Joint Undertaking and several European states, aims to develop sovereign technologies for future european Exascale architectures.

Based on SiPearl's Rhea1 processor. With a budget of €40.8 million over four years, the project brings together European industrial and academic partners, including Atos, CEA, Inria, and GENCI.
Delays affecting the Rhea1 processor, whose tape-out only took place in June 2025, led the consortium to refocus the project toward a stronger software-oriented approach using Grace Hopper and Grace-Grace platforms. Despite reduced hardware ambitions and the discontinuation of GPU integration, EUPEX continued integrating its HPC software stack and optimizing applications for ARM architectures, while maintaining strong collaborative momentum and its vision of European technological sovereignty.

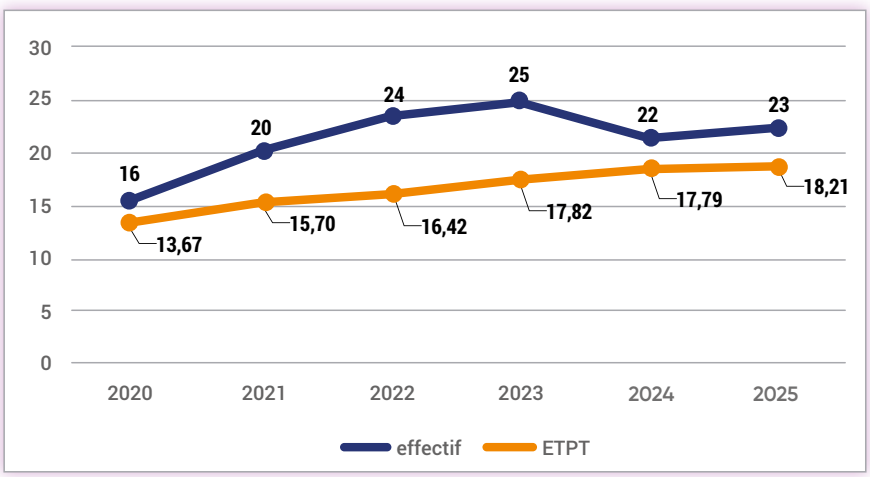
EUPEX presented its work and vision at the HiPEAC conference held in Kraków (Poland) in early February 2026.

Adapting processes to a changed situation

In a context of significant increases in activity, GENCI continued in 2025 to consolidate its internal processes, strengthen its compliance framework and adapt its human resources in order to further secure its operations and effectively support its strategic projects.

Significant work has been carried out to organise the expenditure commitment process with the various budget managers, basing it on delegations of authority that enable GENCI to commit expenditure within a specific framework and at an appropriate level, in accordance with public procurement rules.
This scheme was introduced during the first quarter of 2025. The signing of the delegation agreements was preceded by training sessions for all staff and further specific training for the delegated employees.

The following table present the evolution GENCI's workforce since 2020.



ONGOING IMPROVEMENT OF INTERNAL PROCESSES

Specific training sessions were organized to train and raise staff awareness of the new internal charts that came into force at the end of 2024:

- **the guidelines on best practices for procurement processes,** aimed at ensuring the legal compliance of procurement procedures and guaranteeing adherence to the key principles of public procurement, ensuring the proper use of public funds, promoting responsible procurement (including, in particular, sustainable development and social inclusion) and protecting GENCI and its employees by encouraging the adoption of best practices;
- **the Code of Ethics,** which sets out the rules of professional conduct that staff must adhere to in order to prevent any form of infringement, fraud or offence in their dealings with external partners, and to avoid potential conflicts of interest. It is intended for all GENCI employees, at all levels of responsibility, encouraging them to act with ethics and integrity in the course of their professional activities, whilst upholding GENCI's reputation.

As part of its compliance obligations regarding personal

GENCI, A CIVIL COMPANY

By ensuring compliance, organizing processes and motivating teams, GENCI is strengthening its foundations to provide long-term support for the roll-out of its strategic projects.

EDOUARD BRUNEL, GENERAL SECRETARY AT GENCI

data, in addition to its internal measures for monitoring and updating compliance, GENCI also continues to monitor the allocation of access to computers under the GDPR-compliant subcontracting agreements signed with each of the three data

SOCIAL AND ECONOMIC COMMITTEE (CSE) ELECTIONS

New elections were held in the final quarter of 2025. One permanent seat and one alternate seat were up for election. One permanent employee and one alternate employee were elected. Following a training session for the two elected employees, the first CSE meeting was held in December. Management is delighted with the election results following a four-year period without employee representation, as the establishment of the CSE will enable us to enhance and enrich social dialogue at GENCI.

HR DEVELOPMENTS

Activities related to the Exascale and HQI projects (HPCQS, EUROQCS, EUROHPC-Integration), AI2F, as well as other projects (CLUSSTER, DALIA, ADASTRA, EUPEX) – some of which involve highly complex legal and financial arrangements – have placed a very heavy burden on GENCI's teams in technical, administrative, legal and financial terms.

In addition to these major projects, there are legal activities with the different centres, leading to a significant increase in legal and

GENCI's team in 2025.

In 2025, GENCI continues to monitor the allocation of access to computers under the GDPR-compliant subcontracting agreements signed with each of the three data centres.

financial work: agreements with the centres, particularly regarding the hosting of the exascale machine at the CEA, and associated acceptance testing phases as part of the renewal of the centres' computing and storage resources. These varied and intense activities highlight the consistently heavy workload faced by GENCI's teams. To address this sustained workload, GENCI has launched a study with the support of a service provider specialising in human resources and organisational change management to implement a workload plan across all GENCI projects and activities, thereby improving workload management. The workplan is being developed in parallel with an initiative on Job and Career Path Management. It aims to anticipate, adapt to and support changes in roles, skills and career paths within GENCI.

To manage this workload and address the challenges associated with new projects, GENCI has hired four new staff members: an AI project manager, an AI programme manager, a junior legal officer and a communications intern, for a total workforce of 23.

Internal organization

Management Committee

Philippe LAVOCAT
CEO
(until January 2026 - retirement)

Michaël KRAJECKI
CEO
(since February 2026)

Édouard BRUNEL
General Secretary

Jean-Philippe PROUX
Operations and Security Officer

Stéphane REQUENA
CTO and Innovation Manager

Marie-Hélène VOUETTE
Partnership Manager
Institutional Relationship Advisor

Administrative & Finance department

Maïté CAMPEAS
Executive Assistant

Christèle COCHARD
Office manager

Jason LAFRENIÈRE
Administration & Finance Manager

Marieke PODEVIN
Senior Legal Officer Data Protection Officer (DPO)

Célia LAGIÈRE
Legal Officer

Rasa FALLAHZADEH
Administrative project Manager

Andreea GILBERT
Legal Officer
(since June 2025)

Technical & Operations dept.

Éric BOYER
Project Officer
HPC/HPDA/AI

Cédric AULIAC
AI Program Manager

Raphaëlle ACHACH
AI Project Manager

Anna RIVET
Project Operations Manager

Sabine MEHR
Chief Quantum Project Officer

Corinne BEAL
Exascale Project Manager

Félix GIVOIS
Quantum computing project engineer

Guillaume LECHANTRE
Infrastructure Project Manager

Communication

Nicolas BELOT
Communication Manager and European relations

Annabel TRUONG
Communication Officer

Committees members and official bodies (at 12/31/2025)

MEMBERSHIP OF THE GENCI COUNCIL

State and MESRE representative: Mr Laurent CROUZET - Mr. Ivan MOSZER
CEA representative: Mrs Marie-Hélène MATHON - Mr Hervé DESVAUX
CNRS representative: Mrs Adeline NAZARENKO - Mr Denis VEYNANTE
Universities' representative: Mr Pierre-Alain MULLER - Mr. Olivier SIMONIN
Inria representative: Mr Jean-Yves BERTHOU

MEMBERSHIP OF THE ADMINISTRATIVE AND FINANCIAL COMMITTEE (CCAF)

State and MESRE representative: Mr Antoine PERRANG, Chairman of CCAF
CNRS representative: Mrs Myriam STEPHANUS
CEA representative: Mrs Louise JULLIEN-TAMISIER
Inria representative: Mrs Catherine SUPPER
Universities' representative: Mr Michel DELLACASAGRANDE

MEMBERSHIP OF THE COMMITEE AUDIT CONTRACTS (CM)

State representative:
Mr Philippe AJUELOS, *Ministerial Administrator of Data, Algorithms and Source Codes – digital education department of the Ministry of National Education and Youth, President of the Markets Commission*
Mrs Lara MONTANTIN, *Legal Officer performance, financing and contractualization with research Department of organization management support and regulation*
CEA representative:
Mr Éric STEHLE, *Head of Commercial Service and Strategic Partnership of Paris-Saclay (CEA)*
Mr Thibault PELLETIER, *Head of Reference Service and the Markets Commission at the Commercial and Strategic Partnership Department (CEA)*
CNRS representative:
Mr Sébastien TURCI, *Deputy Director for Purchasing and Innovation of the CNRS*
Mrs Marie-Laure COLIN, *Head of Public Economic Law and Regulations at CNRS*

TECHNICAL ADVISORY COMMITEE (CT)

State and MESRE representative:
Mr Arnault LESERVOT, *Digital Services and Infrastructures Department*
CEA representative:
Mme Valérie BrENNER, *Fundamental Research Department*
Mr Émeric BRUN, *Directorate for Nuclear Energy*
Mr Nicolas LARDJANE, *Head of TGCC*
CNRS representative:
Mr Michael KRAJECKI, *Scientific Delegate*
Mr Pierre-François LAVALLÉE, *Director of IDRIS*
Universities' representative:
Mr François BODIN, *Professor at University Rennes 1*
Mr Michel ROBERT then Michel LABADIE, *Director of CINES*
Inria representative:
Mr Lucas NUSSBAUM, *Head of the Experimental Platforms Program at the Deputy Directorate General for Innovation*

COMMITTEE FOR STRATEGIC GUIDANCE ON INVESTMENTS (COSI)

Committee with responsibility for advising and providing assistance to GENCI in its investment strategy covering in particular the compilation of a multi-year investment plan based on the indications given by the MESRE scientific committee and guiding the work of the CT and CCAF.
State and MESRE representative:
Mr Laurent CROUZET, *Head of Numerical Services and Infrastructures Department*
CEA representative:
Mr Christophe CALVIN, *Fundamental Research Department*
CNRS representative:
Mr Denis VEYNANTE, *Chairman for the Steering Committee of the Compute & Data Mission*
Universities' representative:
Mr Olivier SIMONIN, *Chairman of National Polytechnique Institute (Toulouse)*
Inria representative:
No representative.

GENCI IN 2026

In 2026, GENCI's activities will be marked by numerous challenges.

INFRASTRUCTURES, USERS AND PROJECTS

- → Launch of the Lucy photonic quantum computer
- → Selection of the supplier for the *Alice Recoque* storage system
- → AI Factory France portal
- → Expansion of ADASTRA
- → Acquisition of a cat-type qubit quantum computer
- → Activities of the Maisons du Quantique
- → Calls for projects A20 and A21
- → Scientific results

EVENTS

EuroHPC summit Week 2026

ISC

Forum ORAP

Vivatech

AI Action Summit

France Quantum

Supercomputing

AND SO MUCH MORE...

A huge thank you goes to the authors and contributors for your time and the energy you have dedicated to this Annual Report outlining all the work of GENCI during 2025.

It is only thanks to the people in this report, as well as many others, that GENCI is able to achieve its objectives and be part of the emergence of new possibilities every year.

Publication director: Michaël KRAJECKI – **Coordination:** Annabel TRUONG – **Design & production:** avec des mots – **Photos & illustrations:** Atos ; X. Bouzas ; CEA ; Jean-Pierre Chaboureau ; CINES ; CNRS Photothèque ; DR ; EuroHPC ; Eviden ; France Universités ; Remy Gabalda/MaxPPP ; GENCI ; Pascal Guittet ; IDRIS ; Inria ; iStock ; NVIDIA ; R. Pak/CNRS Photothèque ; PyannoteAI ; Radio France ; Safran ; Nathanael Schaeffer ; TGCC/CEA ; Karine Truffin – **Printing:** Quarante Six – GENCI : 6 bis, rue Auguste Vitu 75015 Paris - France – Tél. : +33 1 42 50 04 15 © GENCI – June 2026

GENCI – 6 bis, rue Auguste Vitu 75015 Paris – France
Tél. : +33 1 42 50 04 15
www.genci.fr

Follow GENCI