



2025 Scientific Report



Foreword

“ Our 2025 Scientific Report attests to the depth and breadth of our research and highlights the most salient results of projects at the cutting edge of innovation.”

ALEXANDRE BOUNOUH
DIRECTOR, CEA-LIST



FOR CEA-LIST, 2025 was a year of exemplary consolidation and forward-looking strategic thinking. With the rise of generative AI and the emergence of physical AI and quantum computing, the pace of technological change has never been faster. In this rapidly-evolving context, our mission of identifying and nurturing breakthroughs in service to responsible innovation has never been timelier.

HCERES, France's higher education and research evaluator, confirmed the excellence of our research in a report that provided a detailed and meaningful analysis of our activities, underscoring our unique position as a bridge between early-stage research and deployable technologies that support our nation's industrial competitiveness. This overwhelmingly positive appraisal positions our research, which includes flourishing contract R&D, among the best in the world. It is on this strong foundation that we have initiated a profound transformation of our programs.

Phase II of our Reveal strategy is at the center of this transformation. Digital continuity, a key enabler of the digital transformation of industry in France and across Europe, is one of the pillars of this ambitious strategy. In January 2026 we launched four major technology programs addressing world models for intelligent robotics, the orchestration of distributed agentic AI, a platform for distributed LLM specialization learning, and multimodal control instrumentation. The goal is to leverage scientific breakthroughs published in A-rank journals and our portfolio of 1850 patents to speed up the development of our technologies and their deployment in real-world industrial use cases.

And these are not the only programs that will support the success of our strategy to speed up the transfer of new technologies to industry. In France, we have deepened our engagement with the CEA (ASIC) and Inria (Numérique) National Program Agencies, a move that allowed us to initiate new programs on high-performance computing components for

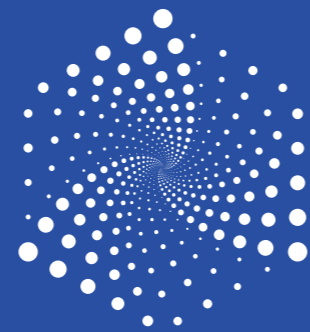
AI (PEPR CAMELIA), component design (PHOENIX), and Digital Twins (EDT). We have also bolstered our commitment to research at the EU level, driving the European vision on robotics and AI and launching the EuroCDP European design platform, whose projects in France will be supported by the CEA-List-led IC-DASH Design Enablement Team.

All our research is conducted with the clear objective of supporting industrial stakeholders through innovation. Among this rich palette of activities, List Tech Days 2025 stood out, bringing in more than 500 decision-makers from industry to discover the latest advances coming out of our labs. This impressive turnout provides further evidence of our central role in the French and European research ecosystems and is a testament to our ability to transform research into tangible innovations for industry.

Our dedication to these programs will shape how we address the challenges ahead in 2026. We have already responded to the changes to the Carnot program and its massive expansion to include, among other initiatives, the French University Innovation Clusters. While the Carnot framework may be changing, its fundamental mission—to stimulate the economy through technological innovation—remains the same. In 2026, Carnot will have to reinvent itself in a more agile, responsive form better able to meet the challenges of our times, from the transition to a green economy to digital sovereignty.

Finally, 2026 will be an opportunity to develop our industrial R&D partnerships through new affiliate programs like DEFI Lab and through expanded use of our technology platforms to support projects with our partners.

Our 2025 Scientific Report attests to the depth and breadth of our research and highlights the most salient results of projects at the cutting edge of innovation. Inside you will find significant advances—spanning all areas of digital technology—that demonstrate our ability to transform early-stage research into concrete innovations that support industrial competitiveness.



Contents

1 Foreword

4 Introduction to the 2025 Scientific Report

6 2025 at a glance

7 Awards and distinctions

10 Initiatives in support of our research

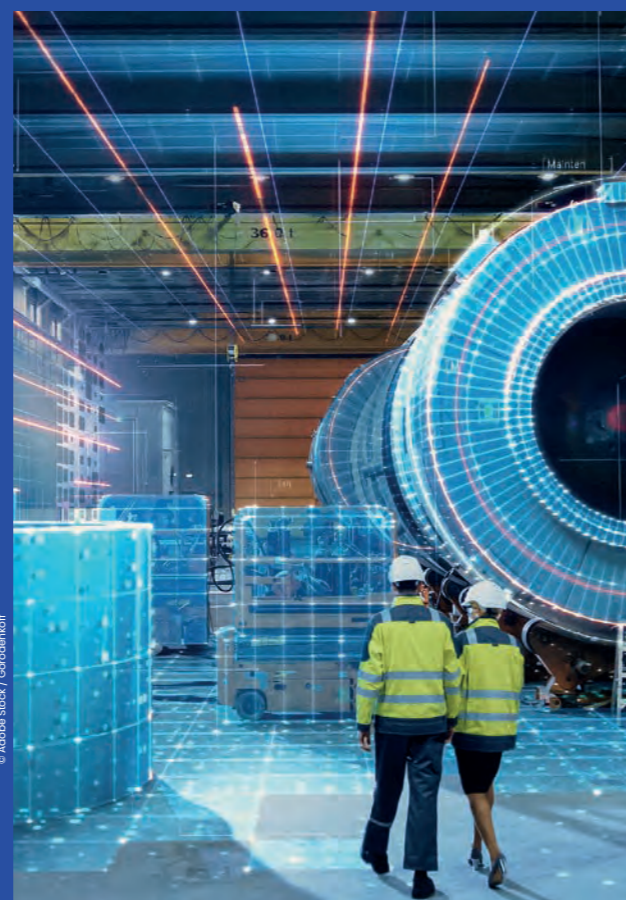
12 2025 highlights

- 12 A European strategy
- 13 Major projects
- 14 Our platforms
- 15 Industrial transfers and applications



16 Systems engineering and Digital Twins

- 18 The Digital Twin, an enabler of flexible, dynamic production lines
- 20 AI for faster multi-physics simulation: an architecture for nonlinear structural mechanics
- 22 Finite element modeling of ultrasonic propagation in viscoelastic media
- 24 Digital Twins: virtual tools with a very real impact
- 26 A machine-learning-based methodology for fast and efficient modeling of power consumption in digital architectures



© Adobe Stock / Gorodenkoff



28 Artificial intelligence technologies

- 30 MiRAG: multi-level retrieval-augmented generation for visual question answering
- 32 Annotation-free world discovery: With xMOD, 2D and 3D vision work together
- 34 3D scene analysis using natural language queries
- 36 Low-latency motion detection with event graphs
- 38 Toward smart acquisition systems
- 40 A federated learning approach with distributed multi-source domain adaptation for serverless collaborative learning
- 42 Adaptive learning to counter catastrophic forgetting and concept drift in federated learning for optimized electric charging station management with predictive capabilities and data privacy
- 44 Analyzing and reducing political bias in large language models
- 46 Verification of neural networks: a challenge to overcome
- 48 Uncertainty in AI-guided Monte Carlo simulations



50 Computing and distributed systems

- 52 Auditing data access with no leaks
- 54 Simulation of attacks on distributed systems
- 56 Making 3D integration a reality
- 58 Embedded non-volatile memory with robust, optimized accesses
- 60 Quantum computing applications and use cases
- 62 Qbricks: toward the formal verification of hybrid classical/quantum programs



64 Digital instrumentation and the factory of the future

- 66 Estimation of a multifingered gripper joint configuration using 3D point clouds and AI
- 68 A cognitive programming interface
- 70 Active lubrication, a kinesthetic and vibrotactile haptic brake
- 72 Fiber Bragg gratings for ultrasonic measurement at high temperatures
- 74 Quantum sensors lift the veil on neutrinos
- 76 Electrochemical measurement techniques for nitrite detection
- 78 Continuous eddy-current monitoring of fatigue degradation in an unnotched steel sample
- 80 AI algorithms for gamma-ray spectrometry used in field measurement
- 82 Photoneutrons and photofission, scientific weapons against trafficking



84 Cybersecurity and trustworthy digital technology

- 86 Incremental software verification for cybersecurity
- 88 Exploitability by attackers and vulnerability prioritization
- 90 Fast fully-homomorphic encryption
- 92 Hardware accelerator tightly coupled to a RISC-V core for post-quantum cryptography

Introduction



CEA-LIST'S world-class research leverages all available collaborative programs in close interaction with academic, institutional, and industrial stakeholders at the national and European levels. Our collaborative research strategy is backed by our investment in and deployment of experimental and R&D platforms equipped with the most advanced technologies. Our goal is to help speed up innovation in service to industry.

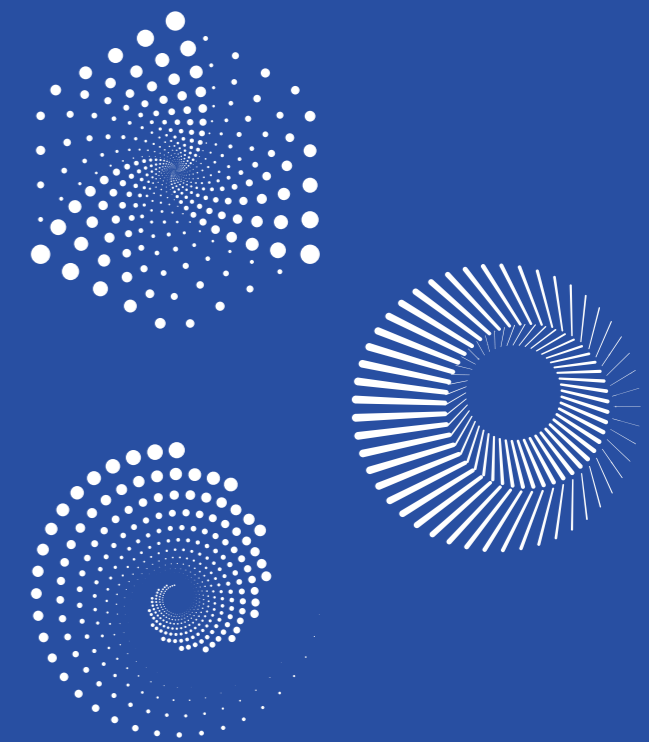
The throughline of all CEA-List's strategic focus areas is to speed up the development of digital technology while maintaining our systems and data engineering expertise at the highest levels, from requirements-based engineering to the development and qualification of complete systems. Central to our approach is the development of software and mathematics-based tools for modeling, simulation, and the construction of Digital Twins. We integrate eco-design principles into all these developments.

Artificial intelligence—which plays a pivotal role in the development of new use cases for computer vision and language and signal processing—is another pillar of our research at CEA-List. Here, we are developing higher-performance, safer, and more frugal and robust approaches for AI designed with integrability into industrial systems in mind. Of course, to deploy digital technology at a large scale and do so sustainably will require appropriate computing solutions, from architectures and components for AI to the Edge-to-cloud continuum to quantum.

Another important area of our research is capturing and processing physical signals and data—the fuel on which digital technology runs. Here, we draw on our long-standing experience in nuclear energy, quality control, and computer perception. We are targeting complete instrumentation and perception chains, from sensors to industrial equipment and systems and from autonomous robots to the factories of the future.

Cybersecurity is, now more than ever, crucial to the development of secure, sovereign technologies. Here, we are focusing on the development of cybersecurity detection, analysis, and qualification tools, technologies, and solutions, as well as on security solutions at the computing architecture and technology levels. Cybersecurity is an integral part of all research at CEA-List.

Our research strategy and organization are the foundation that supports our people in their quest for excellence in research and innovation—excellence they demonstrate every day through their scientific publications and transfers of new technologies to our industrial partners.



© P. Charlet / CEA

© A. Aubert / CEA

2025 At a glance

HEADCOUNT

1000
employees

including

40
state-licensed
research directors

160
PhD students

3
Senior Fellows
(artificial intelligence, computing
architectures, systems engineering)

12
Fellows

63
Senior Experts

113
Experts

36
Research Directors

SCIENTIFIC OUTPUT

300
A-rank
publications

1850
patents

65
new patents
in 2025

TECH TRANSFER

33
startups created
since 2003

137
licensing agreements

200
corporate R&D
partners

RESOURCES

500
projects per year

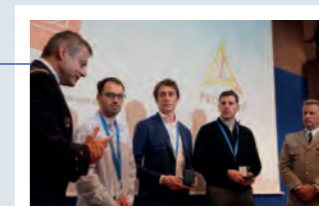
€137M+
annual operating budget

80%+
of our funding is external

Awards and distinctions

The awards and distinctions won during the year are a testament not only to our researchers' scientific excellence, but also to their ability to produce major breakthroughs in their respective fields and to actively contribute to resolving the major scientific, technological, and societal challenges of our time.

- CEA-List software **PyRAT won the AI Safety Challenge** organized by AMIAD, the French Defense Ministry's AI agency, on April 30.
- Our **Colibri software won in the QF_FP (24s) and QF_FPLRA (24s and unsat)** categories for machine reasoning and proof of program at the SMT-COMP'25 competition on August 11 in Glasgow.
- CEA-List tied for **first place for few-shot information extraction and came in second for fine-tuning on closed data** at the EvalLLM Challenge, organized by AMIAD, the French Defense Ministry's AI agency, on June 30 in Marseille.
- The Wavy haptic glove won the **Best Demo Award at the I3E VR 2025 conference** in Saint-Malo in March.
- CEA-List and ASNR (France's nuclear safety authority) won **Euratom's Nuclear Innovation Award** in the Reactor Safety System category for their proposed COMOS monitoring system at the 11th FISA-EURADWASTE conference in May. CEA-List was recognized for its scientific contributions to fiber-optic Raman spectrometry applied to hydrogen containment in severe accidents.
- CEA-List startup **Snowpack won in the i-Nov category of the 2024-2025 National Innovation Competitions** supported by France 2030.
- CEA-List came in second in the Airbus Defense & Space **TextMine'25 competition** for the extraction of document-level relationships between entities from simulated intelligence reports in French. The award was granted at the EGC 2025 conference on January 28 in Strasbourg.



Awards Ceremony for the "AI Safety" Challenge - © COMCYBER



The Wavy haptic glove



Euratom's Nuclear Innovation Award

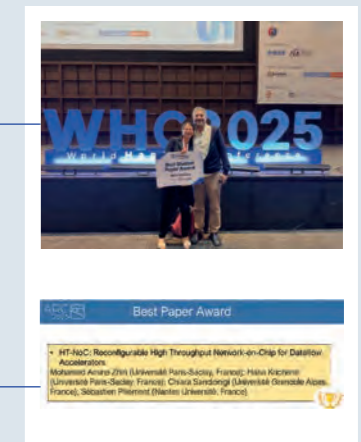


Awards and distinctions

- CEA-List came in **second in the overall rankings at the CheckThat! 2025 Subjectivity Challenge** for a paper entitled “Evaluating LLMs as Detectors of Bias and Opinion in Text” and **came in first for Arabic and Polish**. The awards were granted at the CLEF2025 conference in September in Madrid.
- CEA-List’s PREDIRE sewer inspection robot **won second prize for technology transfer innovations at ERF2025** in March in Stuttgart.
- CEA-List came in **third in the one-shot-method category of the Perception Bin Picking Challenge** at the I3E CVPR conference in June in Nashville.
- CEA-List came in **third in the Welding Quality Detection Challenge** at the Trustworthy AI Summit 2025 in September at EDF Lab in Saclay. The award was given for the automated classification of weld images taken on automotive production lines, an industrial use case provided by Renault.
- CEA-List’s Marielle Malfante was appointed to the **INSITU Chair of Excellence** by Grenoble Alpes University’s MIAI Cluster for artificial intelligence. Her role will be to drive advances in Edge AI and smart sensors for environmental monitoring.
- **Dimitri Kokkonis and Michaël Marcozzi** and their co-authors Emilien Decoux and Stefano Zacchiroli won the **Best Artefact Award** for ROSA, an advanced fuzzing tool that detects backdoors, a type of software vulnerability. The award was granted at the ICSE’25 conference in May in Ottawa.
- A **Notable Review Award** went to **Sébastien Bardin** at the 34th Usenix Security Symposium on computer security and privacy in July in Seattle.
- **Sébastien Bardin** also won a **Distinguished Review Award** at the NDSS Conference in February in San Diego.
- **James Miranda and Ansgar Radermacher** and their co-authors Lukasz Mazur, Nenad Petrovic, Robert Rasche, and Alois Knoll won the **Best Paper Award** for “Querying Large Automotive Software Models: Agentic vs. Direct LLM Approaches” at the GACLM 2025 conference in August in Valencia. The paper explored the application of AI to the management of complex automotive models.



- **Maxence Perion** won the **Best Student Paper Award** at AlgoTel, an international conference on distributed systems, in June in Saint Valéry-sur-Somme for his paper “Stealth Quorums to Counter Adaptive Adversary and Partially Spend Tokens”.
- **Marion Pontreau** won the **Best Student Paper Award** at I3E WHC 2025 (World Haptics) in July in Suwon.
- PhD student **Mohamed Amine Zhiri** won the **Best Paper Award** for his article “HT-NoC: Reconfigurable High Throughput Network-on-Chip for AI Dataflow Accelerators” at ARC 2025 in April in Seville.
- PhD student **Eduardo Tomasi Ribeiro** won the **Best Paper Award** for “HPC Workload Analysis Using Distributed Cross-ISA Binary Instrumentation” at DCIS 2025 in November in Santander.
- PhD student **Lilia Rouzi** won an **Outstanding Paper Award** for “Revisiting Timing Anomalies in Predictable In-Order Pipelines” at ECRTS 2025 in July in Lund.
- **Dorian Lesbre** won the **Doctorant STIC du Plateau de Saclay** (for PhD students in the computer sciences at Saclay) for his research on formal software validation (“Compiling with Abstract Interpretation”).
- The **Best Presentation Award** went to **Aïcha Boukhari** for her presentation on formal software validation at JITA’25 in October in Saclay.
- **Sasha Cuau** won a **Best Poster Award** for her poster on AI robustness at JITA’25 2nd in October in Saclay.
- **Matthieu Lemerre** won the **Best “Speed Poster” Presentation** award at the French National Research Agency’s Interdisciplinary Workshop for Global Security (WISG) in March in Saclay.
- The **Best Poster Award** went to **Abblaihan Uteпов** at ICRM 2025 in May in Paris.



Initiatives in support of our research

AT CEA-LIST we encourage engagement with the academic research community and participation in basic and exploratory research. We believe this is an essential means of supporting our people in their research. France's national Priority Research Programs and Infrastructures (PEPR) were created as part of the government's France 2030 strategy. They focus on medium- and long-term research likely to lead to scientific breakthroughs. What makes these initiatives unique is that, in addition to providing the resources necessary for ambitious research, they are also building national communities around strategic research topics. CEA-List is playing a major role in these national initiatives and is currently involved in nineteen of them in varying capacities, from contributing to or coordinating specific projects to leading entire PEPRs. We are also co-leading (with other large research organizations like CNRS, Inria, and ITM) two PEPRs particularly strategic to CEA-List's research: one on AI and another on cloud computing. We are also either coordinating or contributing to several key projects as part of PEPRs on cybersecurity and on future networks. These projects encourage the formation of new partnerships, but they also foster cooperation between research organizations, which creates opportunities to exchange ideas and identify synergies around topics common to or related to several PEPRs.



NATIONAL PROGRAM AGENCIES

The new National Program Agencies round out the PEPRs by delegating the implementation of research and innovation programs that respond to key economic challenges to leading research organizations. CEA-List is heavily involved in two National Program Agencies, one on architecture, systems, infrastructures, and components (led by the CEA) and one on algorithms, software, and use cases (led by Inria).

In 2025, these National Program Agencies initiated new industry-building programs to support France's industrial competitiveness and sovereignty:

- CAMELIA is addressing high-performance, low-power components for AI. CEA-List and Inria are co-leading this program, which will build on the advances in frugal and Edge AI driven by the PEPR on artificial intelligence.
- Another program, PHOENIX, is developing sovereign systems for heterogeneous high-performance computing. PHOENIX is co-led by CEA-List and CNRS and will leverage microelectronics programs spanning technologies, architectures, and design environments.
- The EDT (Engineering Digital Twins) program will build a systems engineering platform for Digital Twins.
- In the area of AI evaluation and safety, CEA-List is making contributions to AI cybersecurity, the safety of AI-based systems, and controlling bias in generative AI.

FOUNDATIONAL RESEARCH ON ARTIFICIAL INTELLIGENCE: PEPR IA



€73M – 9 separate projects, 5 of which are co-led by CEA-List

200 researchers and 200 PhD students and postdocs recruited nationwide

Priority research topics: Frugal and safe AI; distributed and Edge AI

<https://www.pepr-ia.fr/en/pepr/>



FOUNDATIONAL RESEARCH ON CLOUD COMPUTING: PEPR CLOUD



€51M – 7 separate projects, 2 of which are co-led by CEA-List

192 researchers and 250 PhD students and postdocs recruited nationwide

Priority research topics: Digital continuity, security, virtualization, orchestration

<https://pepr-cloud.fr/en/>



A European strategy

CEA-List is dedicated to making its scientific ecosystem stronger across Europe through an ambitious partnership strategy and a clear positioning.

A particularly active stakeholder in research and innovation, CEA-List is stepping fully into its role, helping to represent the CEA in several key alliances and partnerships strategic to our areas of expertise. These include EFFRA for the factory of the future; ADRA for AI, data and robotics; BDVA for big data; ECSO for cybersecurity; INSIDE for smart digital systems; and Eclipse for open-source software. We are also part of five networks of excellence: AI4Media for AI for the media industry; TAILOR for trust-

worthy AI; HiPeac for computing systems; euRobin for reproducible AI; and dAIEDGE for trustworthy distributed AI.

In addition, CEA-List participates in or coordinates some 100 submissions to a wide range of calls for projects issued by Horizon Europe, Digital Europe, the European Defense Fund, Euratom, EuroHPC, and other EU Joint Undertakings like the Chips JU. The 94 European projects currently in progress address a variety of sectors encompassing industry, digital technology, healthcare, civil security, defense, agriculture, and the environment. We are also tackling societal challenges around the circular economy through projects on digital product passports and the secure

exchange of industrial data at the European level. These projects are, of course, aligned with our eco-innovation program.

Today, Europe needs world-class, high-impact technological research infrastructure where industrial companies of all sizes can test innovative AI-based concepts and products in conditions representative of their production environments. CEA-List leveraged the AI-MATTERS project (one of the European TEFs – Testing and Experimentation Facilities) to create PRISM. This unique facility, through the services it provides to SMEs, is helping make Europe's manufacturing industry more competitive.

Eco-innovation

Environmental responsibility is one of CEA-List's core values. We have been working on eco-innovation for several years now, and our efforts have matured into a research program specifically dedicated to **eco-innovation**. CEA-List researchers published an article detailing how we are helping to reduce the direct environmental and societal impacts of technology (**SustainableTech**) and how we are using technology to help solve environmental and societal problems and assess indirect impacts (**Tech4Sustainability**), as well as the underlying methodologies based on existing standards. Our approach to eco-innovation is based on the conviction that both must be addressed simultaneously if we are to develop technologies backed by an assessment of both positive and negative environmental impacts to arrive at a net benefit.

Our researchers are gradually broadening the use of our eco-innovation process. One example is our research on Digital Product Passports (DPP); another is on software being developed for Life Cycle Assessments (LCA). In the area of waste from electrical and electronic equipment (WEEE), CEA-List technologies are being applied to the recycling of critical materials and the identification of undesirable materials in waste streams.

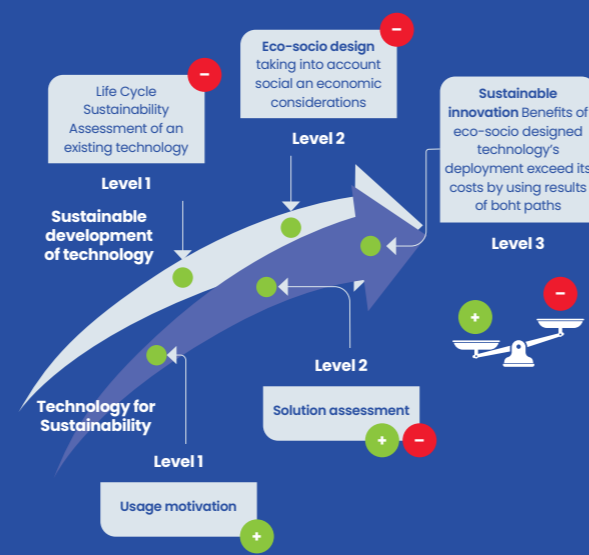


Figure 1: Our eco-innovation process

1 Sustainable technology and Technology for sustainability: The paths towards Eco-innovation, Chiara Sandionigi, Jean-François Berrée, Maxime Peralta, Ariane Piel, Bénédicte Robin, Electronics Goes Green 2024.

Major projects

The people at CEA-List, with their world-class scientific knowledge and deep understanding of industrial culture, have been pivotal in the emergence of ambitious projects to capitalize on and deploy CEA-List innovations.

Here are a few examples:

- The EU STARLIGHT project (2021-2025) built a sustainable, multidisciplinary European community around **AI for security**, bringing together researchers, representatives from law enforcement and industry, plus ethics experts, for a total of 50 partners from eighteen countries. Co-development techniques led to innovations deployed across eight test cases. A total of 50 interoperable AI tools for threat assessment, investigation, and cybersecurity were developed. Furthermore, the project established a robust ethical, legal, and regulatory compliance framework spanning data protection, traceability, and alignment with the AI Act.
- The EU DeepGreen project (deepgreen.ai), backed by a consortium of over 20 partners, was established to build a sovereign platform for the **deployment of embedded AI** open to the entire European industrial ecosystem. The advances made under the DeepGreen project are being integrated into the Aidge open-source platform hosted by the Eclipse Foundation (eclipse.dev/aidge/). The project is helping restore industrial companies' freedom to choose the technologies they use by supporting the development of AI solutions addressing a wide range of established and emerging markets.



© P. Chaillet / CEA

- A series of coordinated projects like EPI and MOSAICS-LP are focusing on the design of **high-performance computing systems** from Edge to cloud, leveraging heterogeneous, chiplet-based computing architectures that can be tailored to specific use cases. The EU CHASSIS project is enabling software-defined vehicles through the development of new architectures and tools.
- IRT Nanoelec's Q-Loop program is addressing the challenge of scaling up **quantum computer control electronics** with error correction. Model-based systems engineering with SysML V2 was used to model a quantum computer system and test it on an attenuation line. The research is now focusing on expanding the models to the entire chain and integrating the visualization and verification of properties.
- The objective of the twelve-partner UDD@Orano project is to integrate advanced digital technologies to speed up nuclear fuel cycle plant modernization. CEA-List helped develop eleven prototypes, hardware, and software to respond to a variety of critical needs, from locating radioactive sources, counting neutrons, and measuring gamma extremity doses to detecting actinide solutions, analyzing complex gamma spectra, and recalibrating simulation models.

Our platforms



© Y. Audic / CEA

CEA-List has created several unique innovation platforms to support our research and facilitate the transfer of new technologies developed in our labs to industry.

These platforms provide structured environments where our engineers, researchers, and industrial partners can work together effectively to develop innovative solutions based on state-of-the-art digital technology. The shared resources—tools, technologies, and expertise—available at these platforms help more rapidly nurture new solutions to higher TRLs and prepare them for deployment in industrial environments.

Dedicated to technologies for the industry of the future, the PRISM platform is home to a unique set of digital resources that can be used for experimentation and testing in near-real-world conditions. Research at PRISM spans robotics and cobotics, Digital Twins, AI and data, software engineering, digital continuity, and additive manufacturing. Once a technology has reached a sufficient TRL, it is integrated into the platform's operational capabilities—the gateway to experimentation on real-world industrial use cases.

CEA-List's innovation platforms are set up to capitalize on and combine technologies, supporting more agile and impactful research. The overriding objective is to speed up the emergence of concrete, sustainable, and competitive solutions to tomorrow's industrial challenges.



© CEA

Industrial transfers and applications

Cultivating both scientific excellence and a solid industrial culture are essential to transforming technological innovations into high-added-value solutions—fulfilling our mission of supporting business' competitiveness. Among the components of this strategy, which produces tangible results in terms of transferring our innovations to industry, are major government-funded research projects at the French and EU levels and the development of top-notch technology platforms. Here is a selection of our most noteworthy advances in 2025:



© Cyril Cortez

SADE and CEA unveiled the TUBOCONTACT SD robot at the Pollutec trade show.

1. SADE: a world-first in robotics

CEA-List and SADE developed TUBOCONTACT SD, an autonomous robot for refurbishing drinking water networks. The robot, which was unveiled at the Pollutec trade show in October, inspects and repairs cast-iron pipes. By eliminating the need to dig trenches, the minimally-invasive robot makes jobs fast, cost-effective, and low- nuisance.

2. AMS software for computer-based interlocking (CBI) control stations for tomorrow's railway signaling systems approved

CEA-List has been working with Compagnie des Signaux since 2021 on the AMS software suite for the specification and formal validation of CBI control stations. In 2025, AMS and its code generation module were approved for CBI control stations by Compagnie des Signaux and French rail infrastructure operator SNCF Réseau.

3. TDK's spintronic memories, a key memory technology for AI systems

TDK and CEA-List renewed their partnership on spin memristors. They will now explore new neuromorphic computing architectures based on this technology, which could enable tomorrow's AI systems to learn and adapt more efficiently with just a hundredth of the power consumption of other components.

4. MBDA Innovation Awards: first place

The CEA has been working with MBDA on embedded AI since 2020. The partners came in first in the 2025 MBDA Innovation Awards in recognition of their joint research.

These successful partnerships are a testament to the positive impact research can have on industry in today's rapidly-changing environment. CEA-List, through sustained innovation and the creation of synergies between academic research and industrial R&D, is helping shape tomorrow's industrial ecosystem.

01

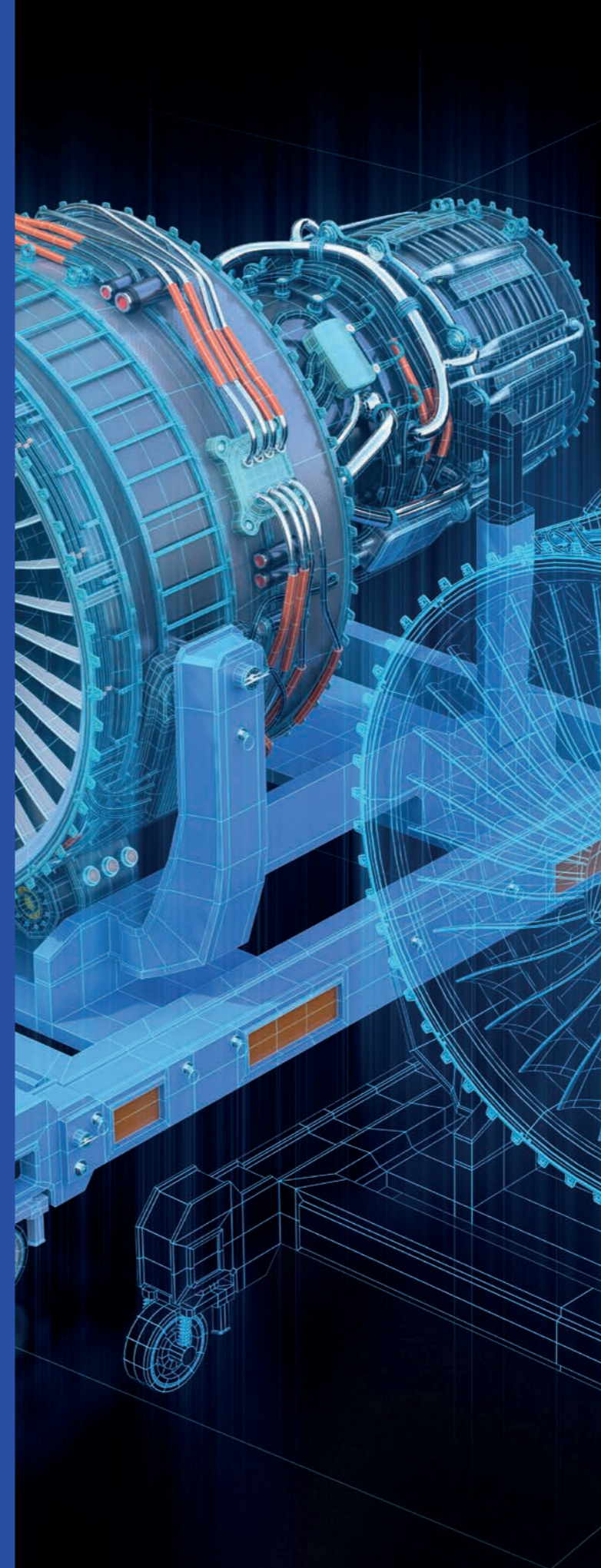
Systems engineering and Digital Twins

Digital Twins are a pillar of systems engineering. By providing decision-assistance capabilities and continuity between the physical and digital worlds, Digital Twins support the transformation of industrial systems. Operational technology (OT) data is gathered and modeled to produce high-fidelity representations (Digital Twins) of real-world systems that can be used to predict unforeseen events and play out the impacts of potential decisions without disrupting the real-world system.

Digital Twins can orchestrate simulation at multiple levels and across multiple business functions, supporting continuous, distributed, tool-based engineering, for more robust, flexible, and safer systems. Looking ahead, dynamic interaction with real-world systems will lead to a controlled and verifiable framework for self-evaluating systems and real-time decision-assistance. Several technological hurdles must still be overcome to bring Digital Twins to their full potential, and CEA-List is addressing these challenges.

Interoperability, for example, will be vital to Digital Twins interfacing seamlessly with existing systems without draining additional resources. One of our solutions is a software architecture that enables both flexibility and automation. Software engineering—which is expensive and requires skilled engineers—is another challenge. We are developing AI-based approaches that speed up physical modeling and simulation while reducing the level of technical expertise required at these stages of development. Finally, we are developing a qualitative methodology to answer crucial questions around the evaluation of environmental and economic costs of massive deployment of Digital Twins in industry.

We are investigating issues like the security of IT/OT communications and the vulnerabilities inherent to tight interactions between the physical and digital worlds; the integration of agentic AI to enable distributed, autonomous, and communicating Digital Twins; and the evaluation of these concepts within environments characterized by shared data spaces as enablers of digital continuity across business functions.



18 Technologies

- • 18 The Digital Twin, an enabler of flexible, dynamic production lines
- • 20 AI for faster multi-physics simulation: an architecture for nonlinear structural mechanics
- • 22 Finite element modeling of ultrasonic propagation in viscoelastic media

24 Methodologies

- • 24 Digital Twins: virtual tools with a very real impact
- • 26 A machine-learning-based methodology for fast and efficient modeling of power consumption in digital architectures

Systems engineering and Digital Twins

Technologies

The Digital Twin, an enabler of flexible, dynamic production lines

On-demand production—one of today’s key challenges in industry—has the capacity to reduce both costs and environmental impacts. CEA-List used its Papyrus4Manufacturing modeling platform to design a new software architecture to provide the flexibility and automation needed to make on-demand production a reality. Tests with CEA-List partner APS confirmed the solution’s potential.

At a time when we are already looking ahead to Industry 5.0, with its focus on people, sustainability, and resilience in production, many of the challenges of Industry 4.0 remain unresolved. One of them is order-driven production. To satisfy a wide variety of orders—whether in terms of number of units or type of product—order-driven production lines must be extremely flexible and automated.

This requires three key capabilities: monitoring production asset availability; dynamically assigning production operations to the assets; and automatically integrating new assets into the existing production process.

Today’s standards and concepts only partially provide these capabilities.

- **The AAS (Asset Administration Shell)** standard for creating Digital Twins of industrial assets includes a management and communication interface.
- **CSS (Capability Skill Services)**, a concept that represents assets (robots, conveyor belts, etc.) by capability, enables the dynamic selection of the most suitable asset for a given job.

- **PnP (Plug-and-Produce)** refers to the ability of an industrial system to be reconfigured, enabling new assets to be integrated into a production process and collaborate automatically and dynamically with other assets.

However, AAS presents certain limitations that prevent it from fully integrating CSS and PnP. For example, AAS cannot detect new assets or provide detailed descriptions of asset capabilities.

CEA-List used its Papyrus4Manufacturing platform, with its AAS-standard Digital Twin modeling and generation tool, to develop XAAS, an extendible AAS software architecture designed to enable flexible, dynamic production lines (Figure 1).

XAAS extends AAS with two major additions:

- **Asset capability description parameters** and
- **A suite of four new services:** Discovery for asset detection, Matchmaking for capacity/job matching, Configuration to configure assets, and Integration for the integration of assets into the current process.



© CEA
QUANG-DUY NGUYEN
Research Engineer



© APS
MAXIME MORIN
R&D Manager, APS

These new features make Plug-and-Produce a reality.

CEA-List implemented the XAAS architecture on a sandbox use case: filling packaging boxes. During the experiment, new assets were automatically integrated into the current production process as soon as they were turned on, and the system correctly assigned tasks to the right robots, confirming the relevance of the XAAS concept.

With XAAS, several asset configurations for a given job can be tested and the best one chosen according to criteria like production time, energy consumption, and the like.

Figure 1: Software architecture of CEA-List’s XAAS solution

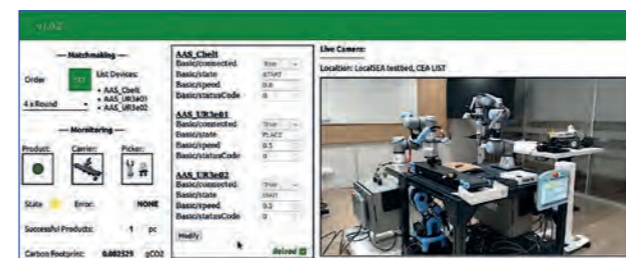
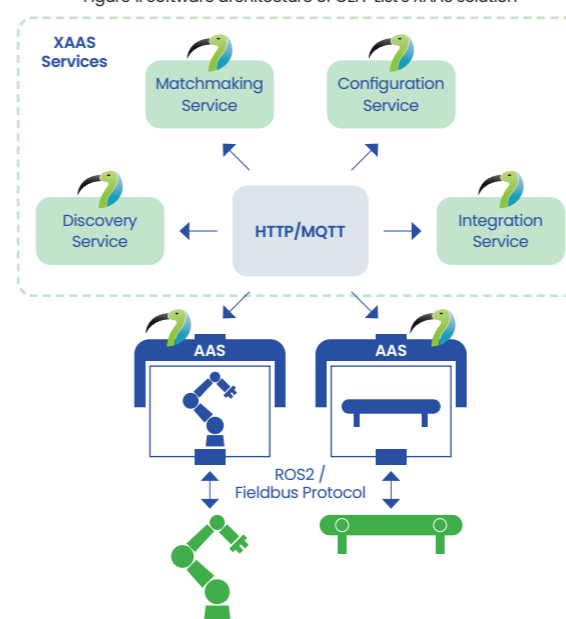


Figure 2: Demonstrator



“ Our Papyrus4Manufacturing modeling platform enables the flexibility and automation that the industry of the future will need.”

••• QUANG-DUY NGUYEN

“ The INTEROP project has confirmed the potential of our architecture using representative production assets. Building on these results, APS has renewed its partnership with CEA-List.”

••• MAXIME MORIN

••• **Prototype**
«Plug&Produce on a production line» - TRL 5.

••• **Industrial project**
Objective: Develop a Plug&Play software component for an assembly line.

••• **Partner**
APS, based in Béziers, France, provides plug-and-play interoperability for all components of production systems.

••• **Flagship publication**
“An AAS-based Architecture for Plug and Produce and Order-Driven Production». Nguyen, Q.-D., Morin, M., Dhoubi, S., & Navarro, N. (2025). *IFAC-PapersOnLine*, 59(10), 927–932. <https://doi.org/10.1016/j.ifacol.2025.09.157>



Systems engineering and Digital Twins

Technologies

AI for faster multi-physics simulation: an architecture for nonlinear structural mechanics

LEBNN (Latent-Energy-Based Neural Network) is a neural network architecture for the mechanics of statically hyperelastic structures. The use of unsupervised learning of the structure's «latent» deformation energy ensures LEBNN's compatibility with coupling schemes, which also means it can be integrated into multi-physics simulations in which some components are simulated by finite elements and others by neural networks.

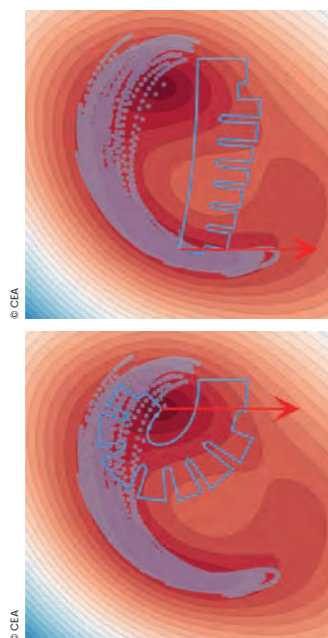
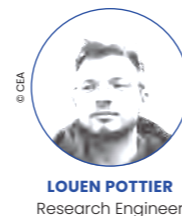


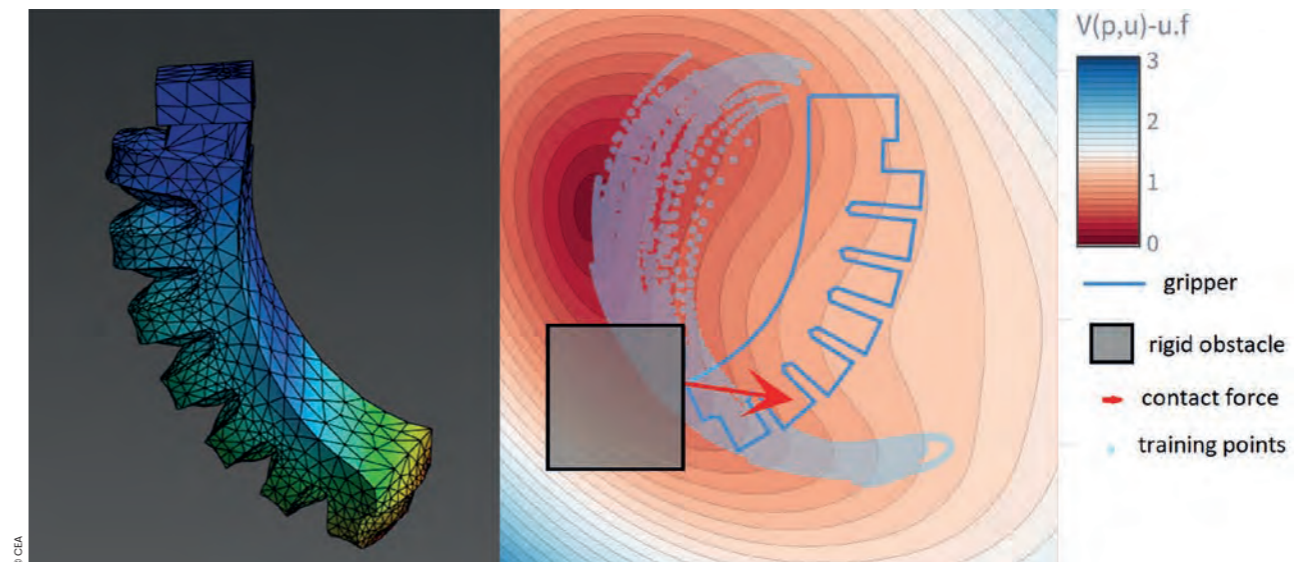
Figure 1: Two displacement fields, u , balanced by the same force, f .

In the statics of non-linear structures, LEBNN solves for the displacement field, u , of a structure subjected to a given force, f . Traditionally, this has been done using the finite element method. When the number of degrees of freedom is large, finite-element solving is too resource-intensive for real-time applications.

Therefore, to speed up the simulation, a dataset of pairs (u, f) can be created and a neural network trained to predict u based on a known f . The goal is to replace the conventional finite-element model

The traditional approach in the literature $u=NN(f)$ has two major shortcomings:

- It cannot learn the behavior of highly non-linear structures with non-convex deformation energy (Figure 1).
- It is impossible to build a coherent rigidity matrix, preventing its use in coupled simulations.

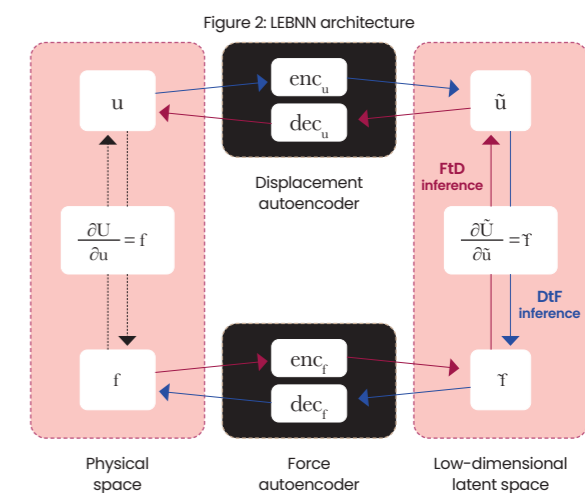


Physical model of the pneumatic gripper used in the ARISE project (left) to train a LEBNN used in a real-time contact simulation (right). The contour lines represent latent energy.

The LEBNN architecture, based on the unsupervised learning of a «latent» deformation energy in a low-dimensional space, solves both problems. Structural properties are encoded in latent deformation energy and, specifically, the topology of this latent deformation energy is the same as the structure's actual deformation energy, which is generally unknown. The structure's rigidity matrix can be calculated from the Hessian of the latent energy. After it is trained, LEBNN can be integrated into a multi-physics simulation, totally replacing the learned structure's finite-element model.

In research for the EU ARISE project, we produced an interactive simulation incorporating a flexible gripper made from hyperelastic materials. Finite-element simulations, which take several minutes, are still a long way from real time. We replaced the gripper's physical model with a LEBNN trained on a hundred simulations. With an inference time under a tenth of a second, coupling with a real-time contact algorithm is now possible.

This demonstration opens the door to the real-time simulation of a new class of systems.



- **Use cases**
Real-time simulation of hyperelastic structures.

- **Major projects** ARISE (EU).

- **Flagship publication**

«Latent-Energy-Based NNs: An interpretable Neural Network architecture for model-order reduction of nonlinear statics in solid mechanics», L. Pottier, A. Thorin, F. Chinesta, *Journal of the Mechanics and Physics of Solids*, Vol. 194, Jan. 2025, Article 105953, pp. 1-13, <https://doi.org/10.1016/j.jmps.2024.105953>

Systems engineering and Digital Twins

Technologies

Finite element modeling of ultrasonic propagation in viscoelastic media

For the simulation of viscoelastic wave propagation, conventional explicit schemes like those used in our CIVA non-destructive testing simulation software become impractical in the presence of localized contrasts or fine meshes emanating from complex geometries. Our research in this area produced two numerical strategies for overcoming this hurdle while maintaining both stability and computational efficiency.

The presence of high-contrast inclusions, thin layers, or highly deformed mesh creates challenges when using standard time-discretization algorithms for the simulation of viscoelastic wave propagation in non-destructive testing (NDT). These features push conventional algorithms to their limits in terms of the maximum time step allowed to ensure correct numerical calculation. The spectral radius of the rigidity operators defined in the «disturbed» area is largely to blame for these standard algorithms' extremely low time steps and disproportionately high computational costs.

We developed and analyzed two numerical strategies for decoupling the overall stability of the time-discretization algorithm from the local rigidity in the «disturbed» zone.

In our first strategy, locally implicit (LI) schemes, formulated by applying an implicit scheme only for local operators, were extended to viscoelastic models. We used an energy method, independent of the operators restricted to the "disturbed" area, to prove the stability of the final algorithm. This strategy was demonstrated on the three viscoelasticity models most commonly used in NDT: Maxwell, Zener, and Kelvin-Voigt. Stability remains governed solely by the background domain—a significant improvement over the explicit reference scheme.

In our second strategy, we extended the explicit scheme stabilization procedure, coupled with a domain decomposition method, to viscoelastic models. In the «disturbed» area, the steep term is replaced by a stabilizing Chebyshev polynomial, the order of which increases the permissible time step proportionally. In practical terms, this is still an explicit method, but one that requires only local resolution at the interface. An improved, controllable stability condition is established by proven energy decay.

We then tested these strategies in realistic 2D and 3D configurations representative of ultrasonic testing in heterogeneous materials. We obtained significant improvements in time steps, number of iterations, and computational cost. The reduction in computational cost was as high as an order of magnitude in some cases, all while faithfully reproducing the attenuation and multiple reflection phenomena looked for in micro-structured materials.



ALEXANDRE IMPERIALE
Research Engineer



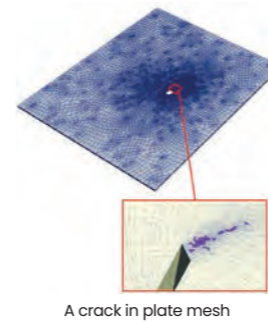
SÉBASTIEN IMPERIALE
Researcher Inria, Ecole Polytechnique, CNRS

« We demonstrated that the limitations of some of the numerical algorithms in our CIVA software when complex geometries and/or materials are involved can be overcome »

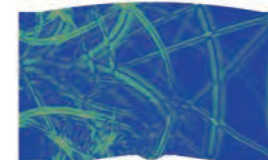
•• ALEXANDRE IMPERIALE

« These methods are a major extension of advanced time-discretization techniques for viscoelastic wave propagation models. »

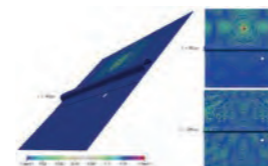
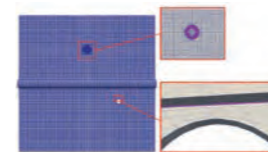
•• SÉBASTIEN IMPERIALE



A crack in plate mesh



Wave propagation in a weld bevel



•• Use cases, applications, technology transfer

Advanced viscoelastic modeling to support the design of high-frequency ultrasonic tests on heterogeneous materials and complex parts (aeronautics, energy, petrochemicals, etc.).

•• Major project and/or partnership

The CARNOT CIVAMONT project, which is funding a post-doctoral contract for Vinuja Vasanthan, whose research focuses on numerical methods for ultrasonic NDT. Collaboration with Sébastien Imperiale, a researcher at Inria.

•• Flagship publication



« Locally implicit and stabilized explicit time schemes for transient visco-elastic wave propagation problems », V. Vasanthan, A. Imperiale, S. Imperiale, *in Journal of Numerical Mathematics*, <https://inria.hal.science/hal-05003766/document>. The paper proposes an extension of the locally implicit and explicit stabilized methods to viscoelastic wave propagation models (Maxwell, Zener, Kelvin-Voigt). 2D and 3D ultrasonic NDT test cases confirmed a significant improvement (in terms of computational cost) over the standard approaches used in the CIVA platform.

Systems engineering and Digital Twins

Methodologies

Digital Twins: virtual tools with a very real impact

Digital Twins, which are built on complex and heterogeneous technologies, are becoming more widely used. The purpose of this research was to come up with a qualitative methodology for assessing the economic and environmental costs associated with Digital Twins. The methodology provides decision-makers with actionable criteria for eco-designing Digital Twins, without the need for more quantitative methods.

Innovations in IoT, AI, virtual reality, and other technologies have positioned Digital Twins as a key technology in recent years. **And yet, the deployment of Digital Twins raises crucial questions about their economic viability and environmental sustainability.**

Industrial maintenance was the first real-world use case for Digital Twins, and it is where the most mature technologies and implementations are currently found. According to our review of the literature, predictive maintenance—enabled by AI and large volumes of data—is the most common use of Digital Twins today. Recent advances in virtual and augmented reality have generated significant additional costs in terms of both hardware and software.

Our qualitative methodology was developed to give industrial maintenance Digital Twins a score indicating their economic and environmental costs. The goal is to help decision-makers ask the right questions from the design stage, when quantitative data is often lacking. We also wanted to propose a tool that is more accessible than more comprehensive methods, like the Life Cycle Assessment (LCA), and that provides a simplified eco-design framework.

Our methodology is based on three key criteria, which, scored on a scale of one to five, determine a Digital Twin's economic and environmental costs (Figure 1):

- **Maturity:** A mature Digital Twin is defined as a Digital Twin with advanced features like prediction and enhanced decision-making, thus requiring more data.
- **Equipment acquisition:** The hardware infrastructure has a direct impact on the environmental costs of Digital Twins. For example, a Digital Twin that uses existing equipment and pre-installed sensors will have a smaller environmental footprint than a system requiring new hardware.
- **User experience:** A Digital Twin with a basic interface, such as dashboards, will have lower economic and environmental impacts than one with virtual or augmented reality.

The scores for each of these three criteria are added up to arrive at an overall score. The higher the score, the greater the environmental and economic impacts. We applied our method to three Digital Twins from the literature: a predictive maintenance Digital Twin (Cognitive DT), a collaborative Digital Twin for remote maintenance (Collaborative DT), and a remaining useful life prediction Digital Twin (DRDT). Figure 2 shows the results of the evaluation of these three Digital Twins.



KENZA ELBAROUDI
Intern
Research Engineer



JEAN-FRANÇOIS BERRÉE
Research Engineer



© Adabi Stock - Zineiron

Although our methodology can be used to score various types of Digital Twins, it does have its limitations. For example, data management is not included, and it does not cover the entire lifecycle. Finally, it doesn't distinguish between economic and environmental costs, which are, in some cases, contradictory.

In its current form, our methodology does provide an initial understanding of the economic and environmental impacts of Digital Twins. However, additional development will be needed to make the tool more reliable.

	Maturity	Equipment	UX
L1	Mirror	No new equipment	Basic visualizations
L2	Shadow	IoT	Advanced visualizations
L3	Control	One simple user terminal	3D modelisation
L4	Cognitive	One augmented user terminal	Extended Reality (XR)
L5	Collaborative	Multiple user terminals	Gamification

Figure 1 - DT assessment grid

« This approach helps decision-makers ask the right questions and provides them with a more accessible and faster starting point for eco-design. »

••• JEAN-FRANÇOIS BERRÉE

80%

Nearly 80% of the environmental impacts of digital products and technologies stem from the equipment itself—from IoT devices to computer displays.

Flagship publication



«Towards A Methodological Framework For Early Qualitative Assessment Of The Ecological And Economic Costs Of Digital Twins In Industrial Maintenance»
<https://doi.org/10.3850/978-981-94-3281-3-ESREL-SRA-E2025-P5071-cd>
 K. Elbaroudi, N. Julien, J-F. Berrée, R. Levailant, A. Piel. *This study was presented at the joint European Safety and Reliability (ESREL) and Society for Risk Analysis Europe (SRA-E) conferences in Norway in June 2025.*



Figure 2 - Assessment of three example DTs

Systems engineering and Digital Twins

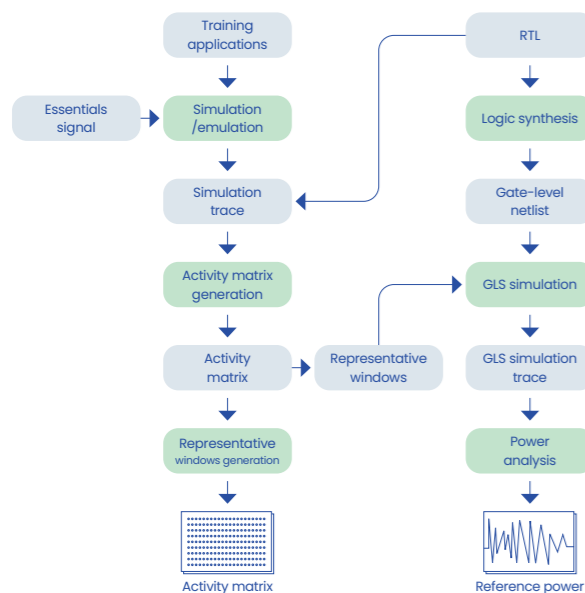
Methodologies

A machine-learning-based methodology for fast and efficient modeling of power consumption in digital architectures



In this research, we proposed an AI-assisted methodology to significantly speed up power consumption modeling. Clustering techniques are used to select representative windows, automatically reducing register transfer level (RTL) traces. This drastically reduces the amount of data and generation times required without sacrificing model accuracy.

Figure 1: Reduced training dataset generation flow.



The purpose of the proposed methodology (Figure 1) is to leverage a combination of machine learning and a drastic reduction in the size of training datasets to speed up power consumption modeling for digital architectures. Traditionally, these models are generated using datasets that contain RTL traces describing the circuit's internal activity and power consumption data at logic-gate level obtained using commercial power analysis tools. However, logic-gate simulations are cumbersome, and generating the corresponding power consumption profiles is costly. Our methodology overcomes these two major bottlenecks.

We used k-means clustering to automate the selection of representative windows. First, RTL traces are segmented into fixed-size windows and then aggregated to enable robust clustering on complex time series.

The analysis is thus restricted to a small but informative subset of runtime segments. The gate-level simulations and corresponding power calculations are based on this subset. The amount of data to process is drastically

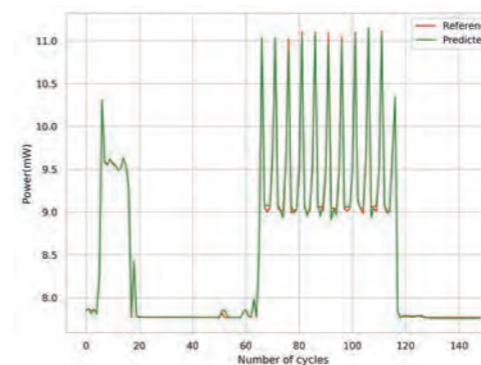


Figure 2: Example of comparison between reference power and predicted power for a masked AES.

reduced, but the behavioral diversity needed for learning is preserved.

Experimental results, obtained on a RISC-V Rocket core and a masked AES, show significant gains: Power consumption analysis time was reduced by up to **28 times** and model training time was reduced by up to **49 times**. The average prediction error was around **5%**, despite the compressed data. As an example, Figure 2 shows a cycle-by-cycle comparison of a segment of predicted power (green) and reference power (red) for a masked AES.

Up to 28x faster data generation and 49x faster training of power models while maintaining a prediction error rate of around 5%.

28x
49x

“ Our AI-assisted approach, which optimizes the selection of training data, significantly reduces simulation requirements without sacrificing the accuracy of power consumption models. This advance will pave the way toward faster, more scalable modeling.”

• CAALIPH ANDRIAMISAINA

• **Use cases, applications, technology transfer**
Automated reduction of representative windows for faster generation of power models for the RISC-V Rocket core. Fast energy analysis in security environments with cryptographic systems like masked AES.

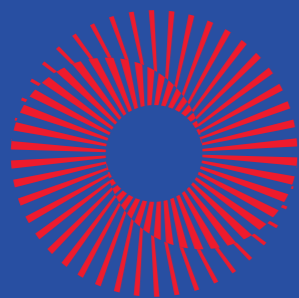
• **Patent**
«Procédé de construction et d'entraînement d'un détecteur de la présence d'anomalies dans un signal temporel, dispositifs et procédé associés». French patent no.: FR2014188 (CEA ref: DD20704)

• **Major projects/partnerships**
This research aligns with CEA-List's AI for EDA activities, which focus on using AI to make hardware architecture design faster and better-optimized. The methodology developed here is currently being implemented as part of the EU TRISTAN project; it will also be used in the EU RIGOLETTO project and in the national CADabrIA project, part of the PHOENIX program. Finally, it is being used in an industrial R&D project with our partner Silvaco.

• **Flagship publication**
“A Methodology for Fast and Efficient ML-based Power Modeling” C. Andriamisaina, K. Trabelsi, P-G. Le Guay, ICCD 2024. <https://doi.org/10.1109/ICCD63220.2024.00109>



02



Artificial intelligence technologies

CEA-List's artificial intelligence research focuses on the development of advanced AI methods spanning generative AI, natural language processing, and computer vision, all in pursuit of robustness, efficiency, and trustworthiness in real-world conditions.

In the field of generative AI, CEA-List developed a beyond-state-of-the-art retrieval-augmented generation (RAG) model for visual question answering about named entities. In computer vision, we are addressing object discovery and segmentation without human annotation through a cross-distillation method that leverages the complementarity of 2D and 3D data, as well as a method for the semantic analysis of 3D scenes from natural language queries for robotics and augmented reality.

In distributed and Edge AI, we are investigating approaches based on asynchronous event graphs that utilize event-driven cameras for very low-latency motion detection. In parallel, near-sensor analysis is driving progress toward smart acquisition systems for structural health monitoring. We are tackling challenges in collaborative and adaptive learning through distributed federated methods for multi-source domain adaptation and through algorithms that limit catastrophic forgetting and concept drift. We are applying these advances to the predictive and private management of electric charging stations, for example.

Finally, the pressing issues of trustworthy and safe AI are top-of-mind in our research on LLM bias analysis, formal verification of neural networks, and uncertainty estimation in AI-assisted Monte Carlo simulations.

30 Generative AI, NLP, vision

- • 30 MiRAG: multi-level retrieval-augmented generation for visual question answering
- • 32 Annotation-free world discovery: With xMOD, 2D and 3D vision work together
- • 34 3D scene analysis using natural language queries

36 Distributed and Edge AI

- • 36 Low-latency motion detection with event graphs
- • 38 Toward smart acquisition systems
- • 40 A federated learning approach with distributed multi-source domain adaptation for serverless collaborative learning
- • 42 Adaptive learning to counter catastrophic forgetting and concept drift in federated learning for optimized electric charging station management with predictive capabilities and data privacy

44 Trustworthy, safe AI

- • 44 Analyzing and reducing political bias in large language models
- • 46 Verification of neural networks: a challenge to overcome
- • 48 Uncertainty in AI-guided Monte Carlo simulations



Artificial intelligence technologies

Generative AI, NLP, vision

MiRAG: multi-level retrieval-augmented generation for visual question answering

CEA-List developed the MiRAG model for visual question answering about named entities. This is the first time a retrieval-augmented generation (RAG)-based approach to generative AI has been applied to this task. MiRAG was demonstrated to be superior to existing approaches in the field based on evaluations on the ViQuAE dataset.



OMAR ADJALI
Post-doctoral student



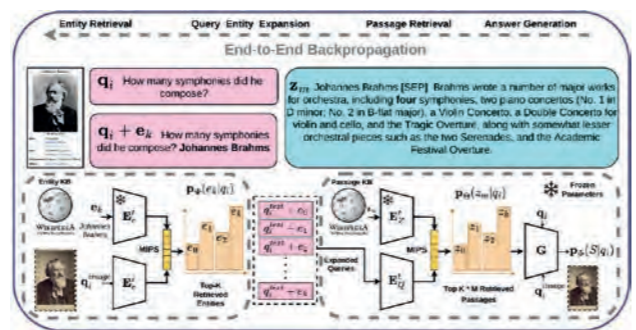
OLIVIER FERRET
Research Director and Fellow Expert



HERVÉ LE BORGNE
Research Director and Senior Expert

Knowledge-based Visual Question Answering about named Entities (KVQAE) has recently garnered attention as a benchmark task for assessing the ability of information retrieval systems to understand visual and textual information in depth. In the standard visual question answering (VQA) task, questions can be answered using images alone. KQVAE, however, requires answers to be retrieved from an external knowledge base made up of both text and images. In a KQVAE system (see figure on the right) the input consists of a question in text form (e.g., "How many symphonies did he compose?") and an image (e.g., a photo of Johannes Brahms). A set of documents combining text and images (Wikipedia pages) must be searched to find the answer.

CEA-List used its RAG-based MiRAG model to complete this KVQAE task, combining an information retrieval model and an answer generation model. The figure above illustrates the main steps in this approach. The basic idea is to search at different levels of granularity (entity and passage) from a multimodal question with textual and visual inputs. Extracting information at different levels gradually refines the question. The answers are then generated using the generation model.



An overview of the MiRAG model.

The search is first carried out at the entity (document) level to identify a set of candidate entities relevant to the question. The entities retrieved are added to the question before a new search is performed at the finer passage level. Adding the entities retrieved during this first step makes the question more specific.

This, in turn, ensures that only the most relevant passages are selected, thus providing the answer generation model with additional context. Stated in more formal terms, the principle is to learn the probability $p(a|q,z,e)$ of generating an answer, a , conditional on

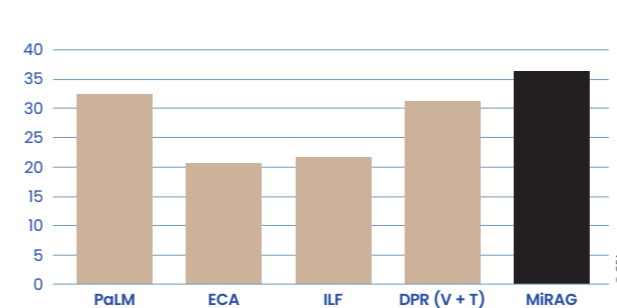


Figure 1: MiRAG results compared with reference methods on the ViQuAE dataset (measurement: exact match).

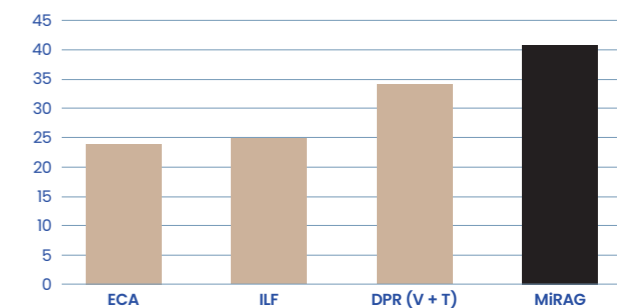


Figure 2: MiRAG results compared with reference methods on the ViQuAE dataset (measurement: F1).

the question, q , an extracted passage, z , and the corresponding extracted entity, e .

Both the information retrieval model and the answer generation model are involved in this end-to-end learning process.

A comparison of MiRAG with reference models combining text and images for the KVQAE task, in this case ECA, ILF, and DPR[V+T], (Figures 1 and 2), shows that combining a RAG-based approach with MiRAG's multi-level search strategy makes better use of the synergy between the two modalities.

While very large text-based generative models like Google's PALM outperform the multimodal models under consideration, MiRAG's use of both text and images significantly outperforms PALM, with 138 times fewer parameters.

Major project and/or partnership

CEA-List developed and tested MiRAG as part of the French National Research Agency MEERQAT project (2020-2024) with the CEA, LISN, IRT and IRISA. The project focused on the multimodal representation (text and image) of entities and the use of these representations for question-answering tasks.

Flagship publication



«Multi-Level Information Retrieval Augmented Generation for Knowledge-based Visual Question Answering».

<https://aclanthology.org/2024.emnlp-main.922/>
Omar Adjali, Olivier Ferret, Sahar Ghannay and Hervé Le Borgne (2024).
In 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2024), p. 16499-16513, Miami, Florida, USA.

Artificial intelligence technologies

Generative AI, NLP, vision

Annotation-free world discovery: With xMOD, 2D and 3D vision work together



One of the major challenges facing computer vision is the human-annotation-free discovery and location of 3D objects. xMOD combines 2D vision (cameras) and 3D vision (LiDAR sensors) in a novel cross-distillation method. The AI learns to segment its environment from motion cues in the images, delivering beyond-state-of-the-art performance.

Conventional object detection requires massive, manually-annotated datasets—making the process costly and difficult to upgrade or scale. We are working on an object-discovery approach that would eliminate the need for human supervision. In 2D image analysis, a group of pixels moving along a coherent trajectory is considered a single object. So far, this approach has not been explored extensively for 3D images, mainly due to the gaps in data inherent to LiDAR point clouds.

Using 2D motion, CEA-List developed an initial reference dataset for 3D object discovery (**DIOD-3D**), and then designed a cross-distillation training framework, **xMOD**, based on a symmetrical “teacher-student” architecture in which a 2D expert model and the 3D model guide each other’s learning. This reciprocal interaction harnesses the complementarity of the 2D and 3D data, compensating for each modality’s weaknesses. The camera steps in for the LiDAR sensor if there are too few data points, and the LiDAR sensor steps in for the camera in challenging conditions like low light or a lack of texture.

The approach also includes another innovation: **scene completion** as a pretext task for 3D discovery. The model

differs from simple reconstruction in that it is trained to predict denser point clouds for a better understanding of object geometries, even when input data is sparse.

With xMOD, millions of scenes can be labeled with no human intervention. The use of both camera and LiDAR data enhances precision.

The model was tested on one synthetic dataset (TRIP-PD) and two real datasets (KITTI, Paymo), outperforming the state of the art in 2D object discovery, with performance improvements ranging from **8.7** (KITTI) to **15** (Waymo) F1 score points. Further reliability improvements can be achieved using a late fusion strategy, which entails combining the 2D and 3D predictions only during inference. This strategy proves particularly effective for objects at medium distances (10 meters to 30 meters).

These advances in object discovery pave the way for faster, more reliable design of the 3D perception models essential to automated driving systems (ADS).

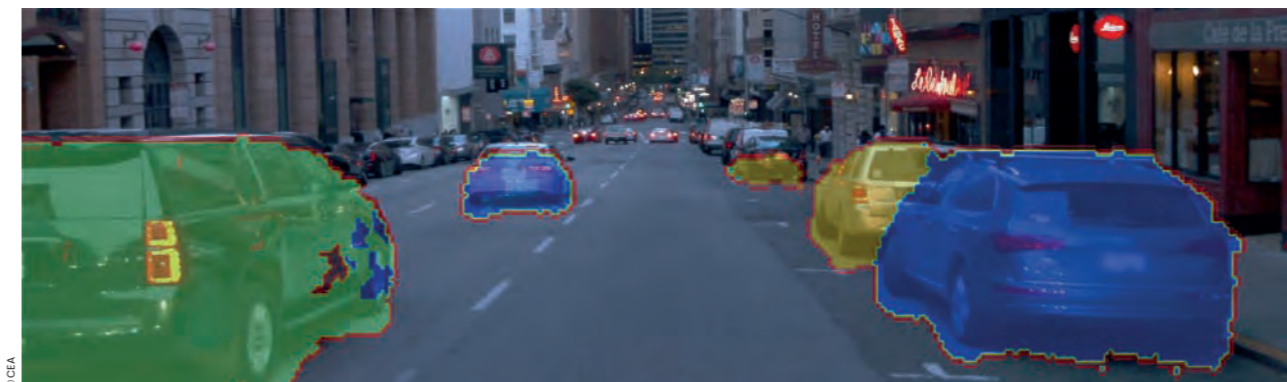


Figure 1: Instance masks predicted by xMOD (2D & 3D)

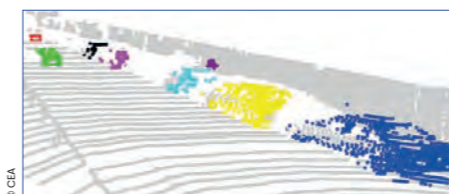


Figure 2: 3D visualization of xMOD predictions (2D & 3D)

« Embedded AI processing into smart systems and devices »

•• SAAD LAHLALI

+15,1

15.1: maximum F1@50 score improvement obtained by xMOD compared to the state-of-the-art in 2D object discovery.

•• Applications
While the algorithm presented here was developed with **autonomous vehicles** and **mobile robotics** in mind, it can be generalized to other systems requiring robust, scalable perception without massive manual labeling. The solution’s redundancy (camera plus LiDAR data) makes it especially well-suited for ensuring robust object discovery in downgraded conditions (low light for cameras; low-reflectivity objects for LiDAR).

•• Project
DeepGreen ANR-23-DEGR-0001. <https://deepgreen.ai/>

•• Flagship publication
«Cross-Modal Distillation for 2D/3D Multi-Object Discovery from 2D motion» <https://arxiv.org/abs/2503.15022> Lahlali et al., CVPR 2025.



Artificial intelligence technologies

Generative AI, NLP, vision

3D scene analysis using natural language queries

The DiSCO-3D semantic segmentation method is used to discover, in a 3D scene, the elements corresponding to the semantic sub-concepts of a user query expressed in natural language. The method, which uses a **NeRF*-based approach**, delivers a high-level understanding of scenes for robotics or augmented reality applications, for example.



The methods currently used for 3D semantic segmentation either identify objects corresponding to a single semantic concept searched for by the user (open vocabulary segmentation, or OV-Seg) or adapt to the content of the scene by discovering several semantic concepts (unsupervised semantic segmentation, or USS).

The broader problem of open vocabulary sub-concept discovery (OV-SD) is effectively addressed by bringing these two paradigms together, for the first time ever, in DiSCO-3D. The objective is to discover the different semantic sub-concepts in the 3D scene that are relevant to a natural language query (Figure 1).



Figure 1: Segmentation of the sub-concepts discovered for the «sleep» (top image) and «furniture» (bottom image) queries.

DiSCO-3D's architecture (Figure 2) is made up of two modules. The first module completes the OV-Seg task to identify areas of the scene that do not correspond to the user's query (the background). The second module "forces" one of the segments to superimpose on the background identified by the first module, creating the USS. This supervision ensures that the other segments discovered by the USS correspond to semantic sub-concepts relevant to the query.

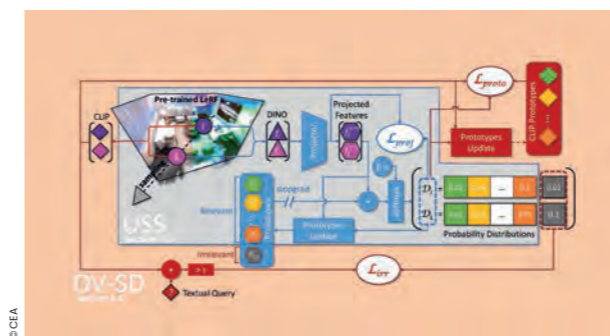


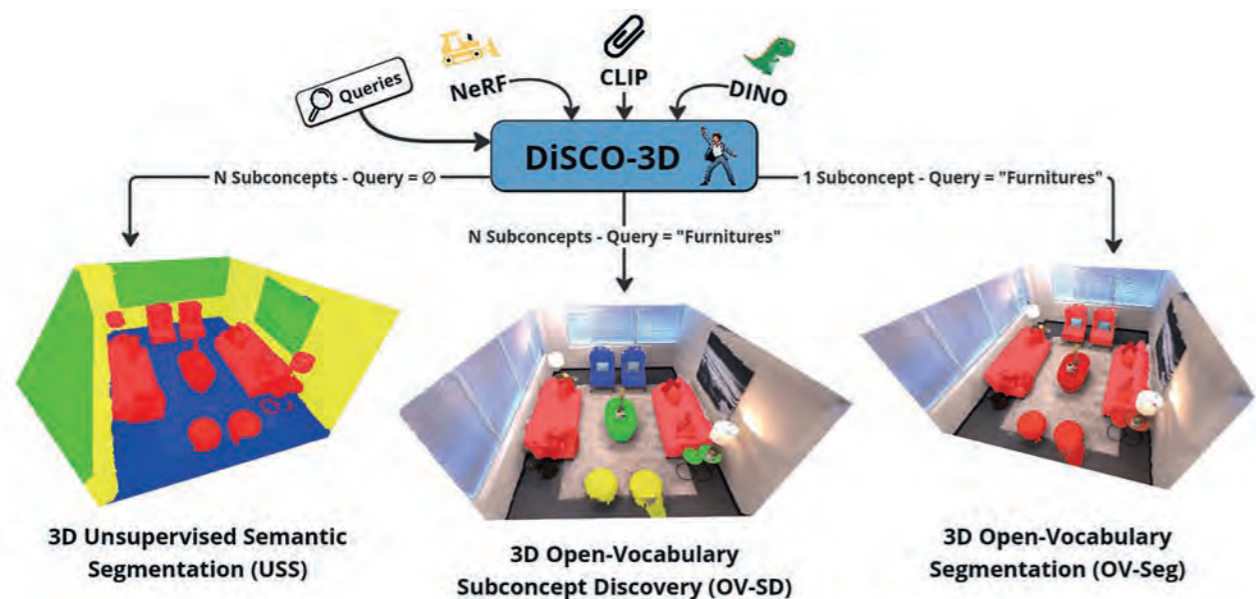
Figure 2: DiSCO-3D for a LeRF** features field

The method proved effective on a variety of user queries on different scenes (Figure 3).



Figure 3: Qualitative evaluation of DiSCO-3D for different queries.

Finally, because the queries are expressed in natural language, DiSCO-3D is easy to integrate as an agentic AI tool. This would make 3D scene analysis using large language models (LLMs) possible.



DiSCO-3D unifies unsupervised and open vocabulary 3D semantic segmentation to discover semantic subconcepts for the content of a specific scene.

- Applications
Reality capture, autonomous robotics.
- Patent
«Méthode de découverte automatique de sous-concepts sémantiques dans une scène», D. Petit, S. Bourgeois; V. Gay-Bellile ; F. Chabot. French patent no. FR2501675.

- Flagship publication
«DiSCO-3D : Discovering and segmenting Sub-Concepts from Open-vocabulary queries in NeRF», Doriand Petit, Steve Bourgeois, Vincent Gay-Bellile, Florian Chabot, Loïc Barthe, The IEEE/CVF International Conference on Computer Vision (ICCV) 2025, Honolulu, Hawaii, USA, pp. 20043-20052 https://openaccess.thecvf.com/content/ICCV2025/html/Petit_DiSCO-3D_Discovering_and_Segmenting_Sub-Concepts_from_Open-vocabulary_Queries_in_ICCV_2025_paper.html

- CEA-List's FactoryIA supercomputer, supported by the Île-de-France Regional Council, made this research possible.

*NeRF: Neural Radiance Fields are a state-of-the-art technology that uses a neural network to reconstruct 3D scenes from 2D images.
**LeRF: Language Embedded Radiance Fields extend the capabilities of NeRFs by associating semantic information with each point in a scene.

Artificial intelligence technologies

Distributed and Edge AI

Low-latency motion detection with event graphs

Event-driven cameras provide low-latency motion detection. Our method leverages asynchronous event graphs that take full advantage of the cameras' high time resolution to **detect motion with very low latency (just 50 milliseconds) while reducing the number of operations 48-fold** compared to the state of the art.



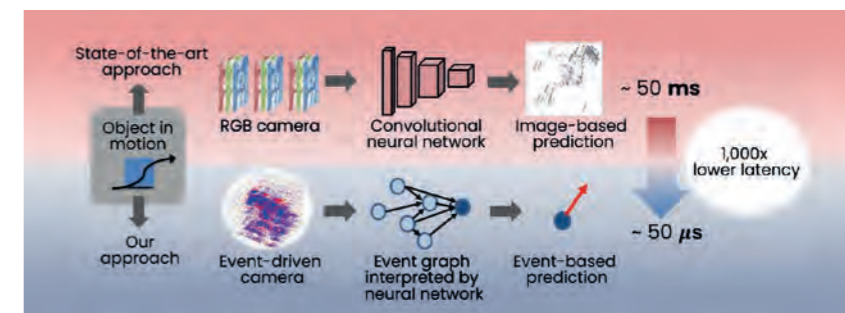
MANON DAMPFHOFFER
Research Engineer



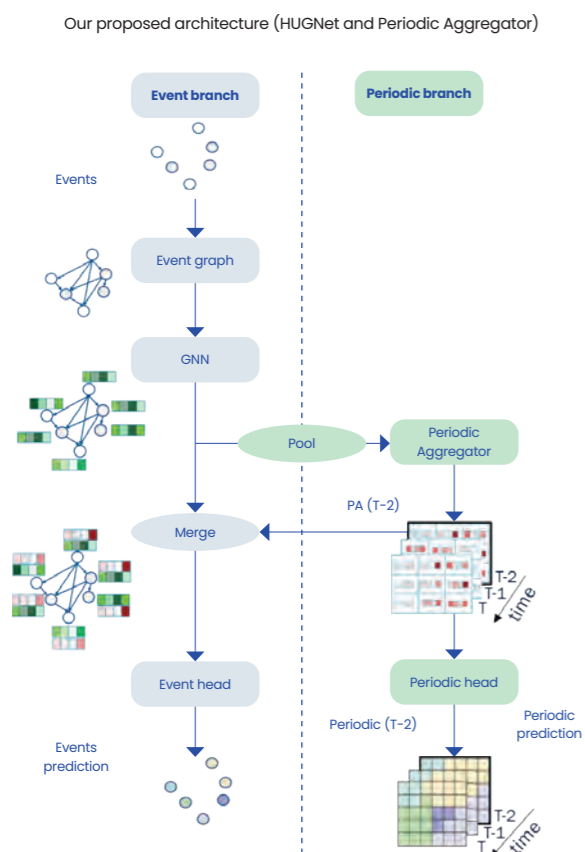
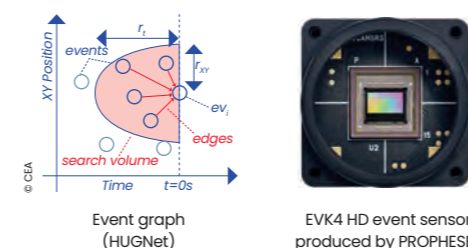
CHRISTOPH POSCH
CTO PROPHESEE

« Our approach marks a step toward low-latency, energy-efficient smart sensors for embedded vision.»

MANON DAMPFHOFFER



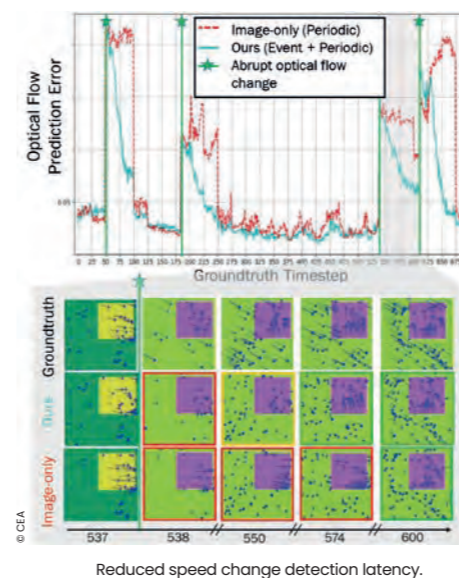
Event-graph-based processing reduces prediction latency by a factor of 1,000 compared to image-based processing.



High temporal accuracy is a prerequisite for computer vision, especially for applications like autonomous driving and video stabilization. Conventional RGB-image-based methods present major limitations in terms of computational cost and image acquisition frequency. With their high time resolution and asynchronous operation, event-driven cameras are a promising alternative. Currently, however, event-driven cameras rely on convolutional neural networks to recreate images from events, cancelling out any latency advantages and making the solution too computationally costly for embedded systems.

Event graphs, interpreted by graph neural networks, enable event-based predictions with minimal latency. The HUGNet method previously developed at CEA-List proved that directed graphs updated on the fly could further reduce latency. However, due to a lack of global context aggregation, the lower latency comes at the cost of lower accuracy. We came up with a new neural network architecture combining the asynchronous branch with a periodic aggregation branch—effectively overcoming this limitation. Using a recurrent convolutional neural network on images of past events, the periodic aggregation branch extracts a global context from the scene. This improves the quality of the predictions without sacrificing latency.

Our approach turned out to be **1,000 times faster than conventional solutions**, reducing speed change detection latency down to a prediction-per-event time of just **50 μs** on embedded hardware. In addition, it requires **17 times fewer parameters** and **48 times fewer operations per second**, while maintaining competitive accuracy levels.



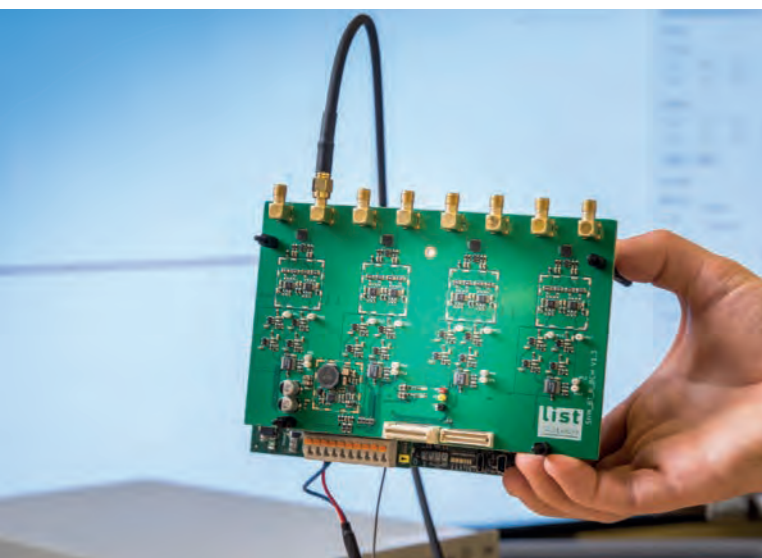
- • > **Use cases, applications, technology transfer**
The algorithm will be integrated into a very-low-power (<10 mW) three-layer smart event imager manufactured in partnership with PROPHESEE and STMicroelectronics. In addition to optical flow prediction, the algorithm can automatically recognize patterns like gestures or lip-reading.
- • > **Patent**
DD24590ST - Pending.
DD22912ST - Filed.
- • > **Major project and/or partnership**
ANR IRT Nanoelec (Smart Imager program), in partnership with PROPHESEE.
- • > **Flagship publication**
«Graph Neural Network Combining Event Stream and Periodic Aggregation for Low-Latency Event-Based Vision».
M. Dampfhofer, T. Mesquida, D. Joubert, T. Dalgaty, P. Vivet and C. Posch (2025).
In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
<https://doi.org/10.1109/CVPR52734.2025.00648>
Selected for the conference «Highlights» (A-rank').

Artificial intelligence technologies

Distributed and Edge AI

Toward smart acquisition systems

Advances in Edge computing and AI have made integrating analysis capabilities directly into acquisition systems a reality. These technologies can aid with non-destructive testing (NDT) and, if structures are instrumented with sensors, automate structural health monitoring (SHM).



Analog acquisition circuit board integrated into the GERONIMO system.

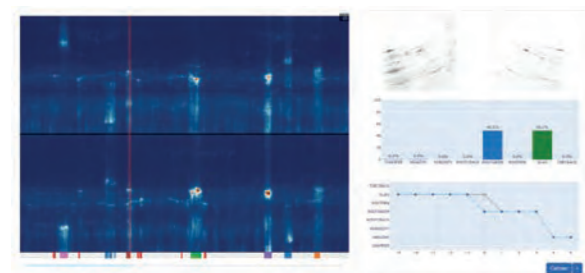


Figure 1: Left: complete inspection of a weld using the two probes (top view). Right: TFM images from both probes, results of the current prediction, and a comparison of the predictions with the expert reference.

Smart acquisition systems are the way of the future. However, what these systems will look like will depend on the use case. For non-destructive testing, for example, the purpose of these systems is to guide inspection and provide decision-assistance to support human operators. In structural health monitoring, sensors generate a continuous and very large stream of data. The automated in-sensor analysis of this data would limit the amount of data exchanged—overcoming a major challenge. At CEA-List, we possess broad, deep knowledge of the hardware used in acquisition systems and have the capacity to integrate AI models and processing directly into sensors. For example, our Aidge framework was developed to manipulate, convert, and optimize neural networks for Edge devices, while our GERONIMO acquisition system for guided-wave measurement enables the deployment of end-to-end solutions, from sensor to diagnosis. CEA-List has developed two proof-of-concept prototypes that illustrate these capabilities.

First, a ConvNeXt-type model, reduced for constrained hardware environments, was implemented on TFM image analysis for ultrasonic weld inspection^[1]. The model, trained on experimental data, classifies images with no false negatives—a non-negotiable in industrial environments—and identifies 87% of healthy data under nominal inspection conditions. To ensure the model is robust, we modified the inspection conditions by moving the probes during inspection and by changing the geometry (pipes, plates, variations in thickness) of the structure inspected. We used Aidge to optimize the model, which was then efficiently deployed on a NXP i.MX8 architecture (Figure 1), the foundation of our different acquisition systems.



CLÉMENT FISHER
Research Engineer



ROBIN GUYON
Research Engineer

« Embedded guided-wave analysis transforms measurement data into clear indicators of the structure's condition, enabling continuous monitoring.»

•• ROBIN GUYON

Embedded data compression: up to 10,000x less data (3 floating values instead of several kB of data)

÷10000

A frequency of around 10 Hz was achieved during inference. In guided-wave structural health monitoring^[2], the goal is to automate the entire chain, from acquisition to decision-making. In this demonstration, we used an image-processing convolutional neural network for delay-and-sum (DAS, Figure 2) image analysis. This qualitative imaging method is not the most efficient in terms of resolution, but it is still widely used at the state of the art and, therefore, remains a relevant test case. Data simulated with CIVA software was used to train the model, which was then adjusted by transfer learning on a few experimental measurements, enabling defects to be located and measured. The model, converted using Aidge and embedded in the GERONIMO acquisition system (Figure 3), reduces the volume of sensor data from several kilobytes to just three floating values. This compression considerably reduces data transmission, enabling continuous monitoring, even on heavily instrumented structures.

We are now working to expand the range of processing methods available to include guided-wave tomography. We are also integrating AI into GERONIMO and its OPTOGERO version for fiber Bragg grating measurement^[3].

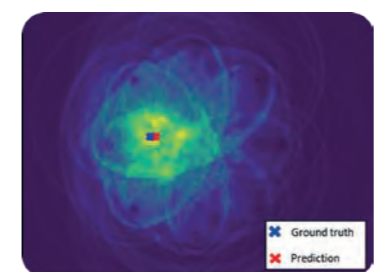


Figure 2: DAS imaging with the model's prediction of the defect's position.

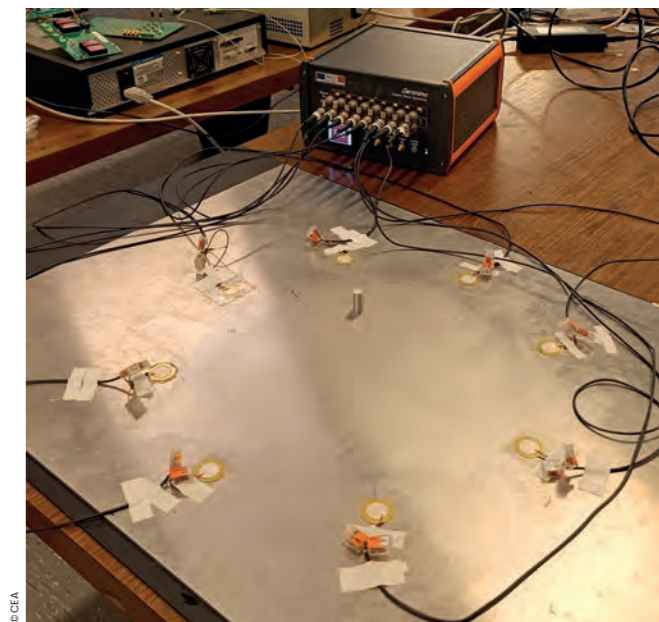


Figure 3: Plate instrumented with the GERONIMO system.

•• Use cases, applications, technology transfer

GERONIMO: Partnership with UGE and licensing agreement with Capturia for commercialization.

•• Major project and/or partnership

Find, Renaissance Fusion, Alstom RSHM, DeepGreen, Multimod'AIR, etc.

•• Flagship publications

[1] «Embedded Artificial Intelligence in Guided Wave SHM system: Signal processing, and data analysis». C. Fisher, A. Recoquillay, B. Chapuis, and P. Calmon, *e-Journal of Nondestructive Testing*, vol. 30, no. 06, Jun. 2025, <https://doi.org/10.58286/31305>

[2] «Towards embedded AI models for welding defect detection in pipes» R. Guyon, M. Newson, C. Fisher, R. Miorelli, and D. Roué, in *2025 IEEE Sensors Applications Symposium (SAS)*, Jul. 2025, pp. 1–6, <https://doi.org/10.1109/SAS65169.2025.11105144>

[3] «Embedded passive guided wave tomography. Application to corrosion monitoring in multilayered pipes» A. Recoquillay et al., *e-Journal of Nondestructive Testing*, vol. 29, no. 07, Jul. 2024, <https://dx.doi.org/10.58286/29855>

Artificial intelligence technologies

Distributed and Edge AI

A federated learning approach with distributed multi-source domain adaptation for serverless collaborative learning

CEA-List developed De-FedDaDiL, a fully distributed method for multi-source domain adaptation (MSDA). Based on Wasserstein barycenters and dataset dictionary learning, De-FedDaDiL enables models to be adapted to an unlabeled target domain without a central server, all while improving robustness, scalability, and privacy.



Understanding the challenge: adapting to heterogeneous data from multiple sources

Multi-source domain adaptation (MSDA) addresses the key challenge of enabling a model to utilize multiple heterogeneous data sources to generalize to an unannotated target domain. In many industrial applications, not only does data vary from location to location, but it is also often impossible to centralize. This creates a distribution mismatch between source and target that severely impedes the performance of conventional models, which are generally based on an assumption of homogeneous data that is rarely found in real-world situations.

The limitations of conventional federated learning: a central server that weakens the entire system

Recent approaches using federated learning have attempted to adapt MSDA to the distributed context, while protecting data privacy. However, virtually all these approaches involve the use of a central server to aggregate and synchronize local models, introducing weaknesses like dependence on a single point, increased cybersecurity risk, and high data transmission costs.

- Use cases**
 This innovation could be useful in inter-institutional collaboration without a central server, the rapid adaptation of models to new domains, and peer-to-peer learning in sensitive environments. This research was funded by the PEPR IA REDEEM project (ANR-23-PEIA-005).

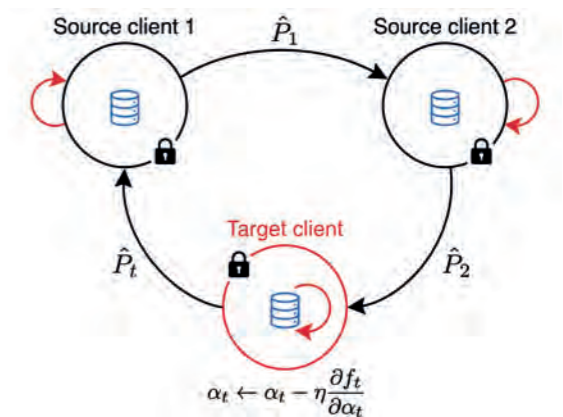
- Flagship publication**

«Decentralized Federated Dataset Dictionary Learning for Multi-Source Domain Adaptation»,
 R. Clain, E. F. Montesuma and F. N. Mboula,
ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Hyderabad, India, 2025, pp. 1-5,
<https://doi.org/10.48550/arXiv.2503.17683>

Our innovation: totally peer-to-peer learning

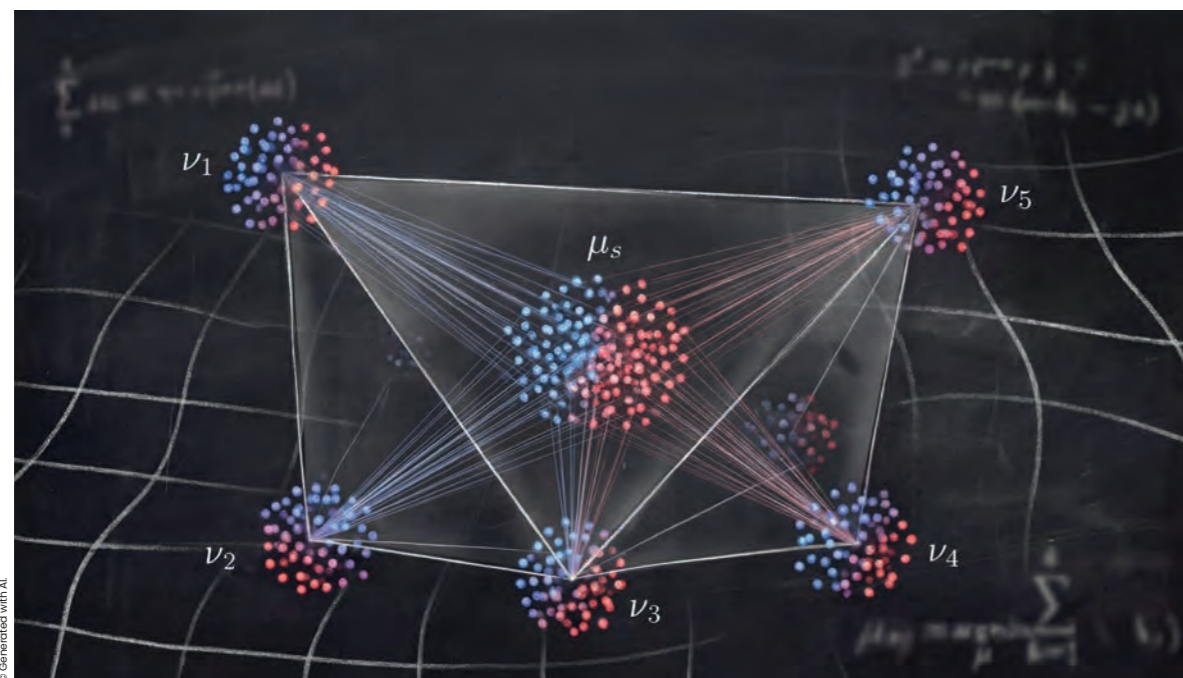
De-FedDaDiL eliminates the central server. Each client stores its own dictionary of atoms (the fundamental elements used to model distributions) locally and exchanges this information with a randomly-chosen peer at regular intervals. The atoms received are merged with those of the client via a simple but effective aggregation process and then optimized locally. The barycentric coordinates, which encode how each domain projects onto the atoms, are never shared, guaranteeing enhanced privacy (Figure 1).

Figure 1: Peer-to-peer exchange of distribution envelopes



Performance that rivals the best federated methods

De-FedDaDiL was tested on the ImageCLEF, Office-31, and Office-Home datasets, where it performed on par with (and, in some cases, better than) the centralized federated version. On ImageCLEF, it even outperformed all the methods tested. The results of these tests indicate a 2% reduction in performance compared to server-based approaches, while cutting the number of data exchanges in half. An analysis of the convergence between clients showed a gradual decrease in the Wasserstein distance between barycenters. This confirms natural alignment despite the totally peer-to-peer approach with no central server.



Barycentric interpolation of a dictionary of datasets in optimal transport geometry.

© Generated with AI

Artificial intelligence technologies

Distributed and Edge AI

Adaptive learning to counter catastrophic forgetting and concept drift in federated learning for optimized electric charging station management with predictive capabilities and data privacy



CEA-List developed federated learning algorithms to predict charging station occupancy in real time without sharing sensitive user data. We demonstrated that our approach could reduce prediction errors by as much as 40% by detecting changes in behavior and limiting model forgetting during updates.

As the energy transition gains traction, the use of distributed energy resources like renewables, electric vehicles, IoT devices, and microgrids is increasing rapidly—multiplying the number of distributed data sources and generating massive amounts of heterogeneous and sometimes-sensitive data. When it comes to processing all this data, centralized data management approaches are reaching their limits in terms of security and privacy. The purpose of the AI-NRGY project (PEPR TASE) is to develop a distributed AI architecture suitable for the very large number of decentralized sources of data and energy that will make up the energy systems of the future. In federated learning (Figure 1), models can be trained collaboratively on distributed data without moving the data, improving data security and privacy. However, when it comes to the processing of time data and continuous data streams, federated learning does present two drawbacks: concept drift (changes in the statistical properties of data over time) and catastrophic forgetting (the loss of previously acquired knowledge when a model is retrained on new data). Our continuous, adaptive federated learning framework efficiently manages concept drift, catastrophic forgetting, and heterogeneous client data.

The process is broken down into two phases, one offline and one online (Figure 2). During the offline phase, data is collected, prepared, and pre-processed at each client site. The backbone federated learning model is then trained. During the online phase, data streams are processed in real time, and the trained federated learning model makes predictions. Model performance is evaluated after each prediction. A drift-detection mechanism identifies any changes in data distribution.

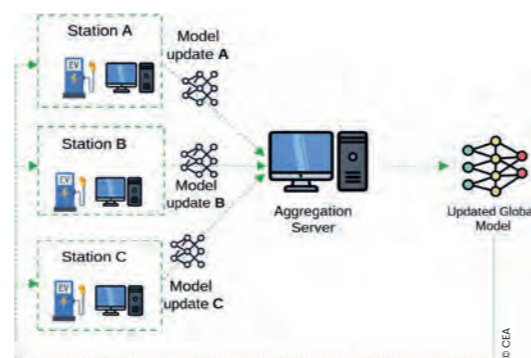


Figure 1.

Up to a 30% to 40% reduction in prediction errors

-30%
-40%

If any drift is detected, adaptation techniques based on dynamic clustering and incremental learning are applied to adjust the model to the changing data. Empirical testing demonstrated a significant reduction in prediction errors (up to 30% to 40% compared to reference models) in the presence of concept drift.

We are developing a method to predict charging station occupancy, improving turnover and reducing wait times. Using federated learning, each charging station is trained on its own data, which is never shared. Only the model parameters are exchanged. Changes in behaviors (seasonal, new user habits, new neighborhoods, etc.) are detected and used to adapt the model without «forgetting» what has already been learned. This solution provides more reliable forecasts for better travel planning and more efficient charging network management—and it protects the privacy of user data.

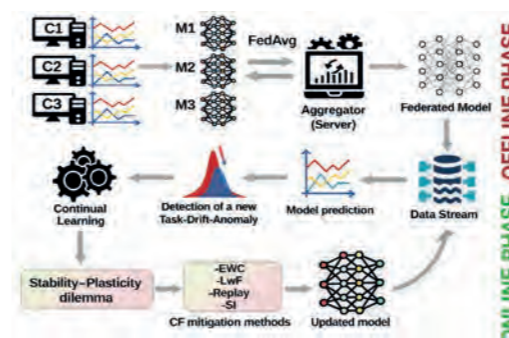


Figure 2.

• • • Application
Decentralized time series prediction.

• • • Major project
Project AI-NRGY / PEPR TASE.

• • • Flagship publication
«Adaptive federated learning framework for predicting EV charging stations occupancy». Hallak, K., and Kem, O. (2025). *International Journal of Transportation Science and Technology*. <https://doi.org/10.1016/j.ijtst.2025.04.007>



Artificial intelligence technologies

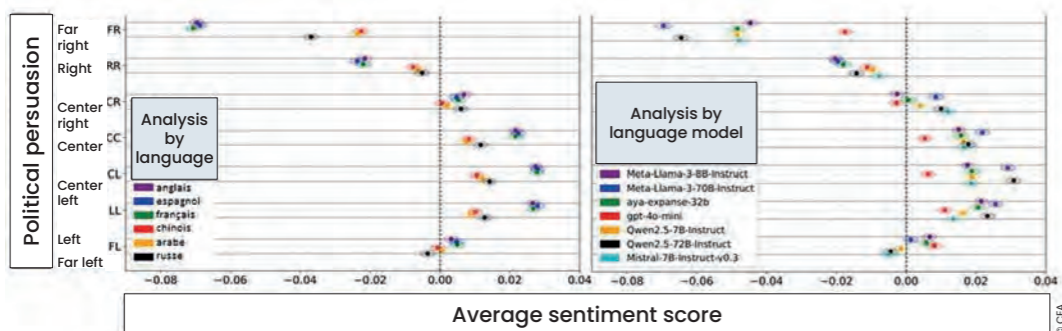
Trustworthy, safe AI

Analyzing and reducing political bias in large language models



AKRAM ELBOUANANI
PhD student

CEA-List developed a new method for measuring political bias in large language models. The results of our research revealed some surprising ideological preferences, which vary according to language and model size. We demonstrated that, when politicians' names are changed, bias is significantly reduced.



Average sentiment by language (left) and aggregated model by politics (right).

CEA-List developed a novel method for analyzing the political biases present in large language models (LLMs). This issue is usually approached by having LLMs answer political questionnaires or generate text—methods that have their limitations. Here, we tackled the problem from a simple, yet effective angle: We replaced the name of one politician with another in identical sentences and then observed how the models' opinions changed.

A sentence on a political topic was chosen from the press, and the model was asked to indicate whether the sentiment was positive, negative, or neutral toward the politician mentioned. The real politician's name was replaced with the names of more than 1,300 other politicians from different regions and of different political persuasions. This was repeated in six languages on seven models, for nearly 25 million predictions. The large amount of data was used to identify systematic variations in the answers produced. We interpreted these variations as forms of bias.

Several clear trends emerged from the results. First, the models included in the experiment (ChatGPT, Qwen, Llama, Aya, and Mistral) demonstrated overall preferences: They more frequently associated a positive sentiment with left- and center-leaning politicians, and a negative tone with right-wing and far-right politicians. These biases appeared in all languages, but were more pronounced in Western languages like English, French, and Spanish.

We also learned something very interesting about the influence of model size: The larger, generally better-performing models also expressed biases more consistently and emphatically. This behavior can be seen in the figure above, which compares the two versions of Llama and Qwen. This insight suggests that not only do the most powerful models produce the most consistent answers, but they also amplify the trends in their training data.

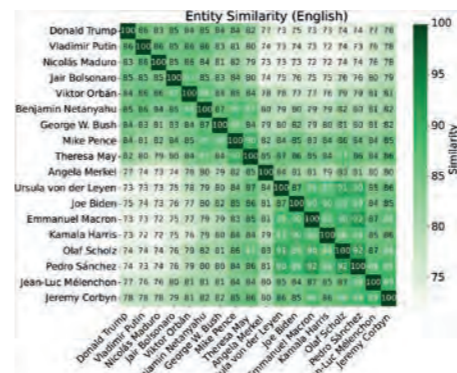


Figure 1. Similarity between political entities

We also demonstrated that the models have internal representations of political figures (Figure 1). A comparison of the predictions associated with different politicians highlighted strong similarities between certain ideologically-similar figures (e.g., between European Social Democrats). This suggests that the models have their own mental organization of the political landscape.

Finally, we tested a method for reducing bias: We simply replaced politicians' real names with fictitious but believable names, while retaining characteristics like gender and country of origin. Although some variations did remain, our test significantly reduced unjustified variations in predictions (Figure 2).

This simple, reproducible method provides a new framework for studying bias in large language models. Our experiments show that even very powerful language models are not immune to the political associations present in their data and suggest that even slight changes to how a prompt is expressed can make sentiment analyses more equitable when dealing with sensitive topics.

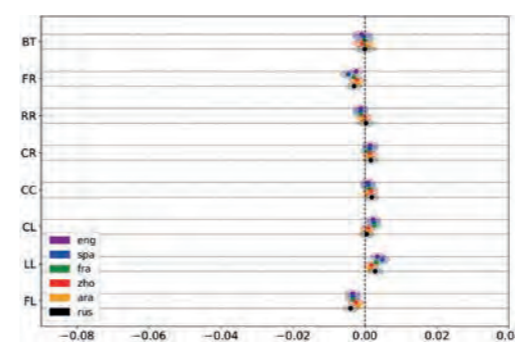


Figure 2. Average sentiment score by language for the political trends studied after bias reduction by anonymization

“Our research shows that not all politicians are treated equally by large language models. In fact, we revealed some very real biases in these models' predictions.”

AKRAM ELBOUANANI

X2

Political bias in LLM sentiment analysis can be cut in half without affecting performance by anonymizing politicians' names.

CEA-List's FactoryIA supercomputer, supported by the Île-de-France Regional Council, made this research possible.

Major projects
BOOM ANR-ANR-20-CE23-0024 and OpenLLM-France.

Flagship publication
“Analyzing Political Bias in LLMs via Target-Oriented Sentiment Classification”. Akram Elbouanani, Evan Dufraisse, and Adrian Popescu. 2025. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 15476–15505, Vienna, Austria. Association for Computational Linguistics. <https://arxiv.org/html/2505.19776v1>
This research is part of a long-term project on algorithmic fairness that addresses the entire AI lifecycle, from building training corpora to monitoring biases in production. The research was published in the ACL 2025 proceedings.

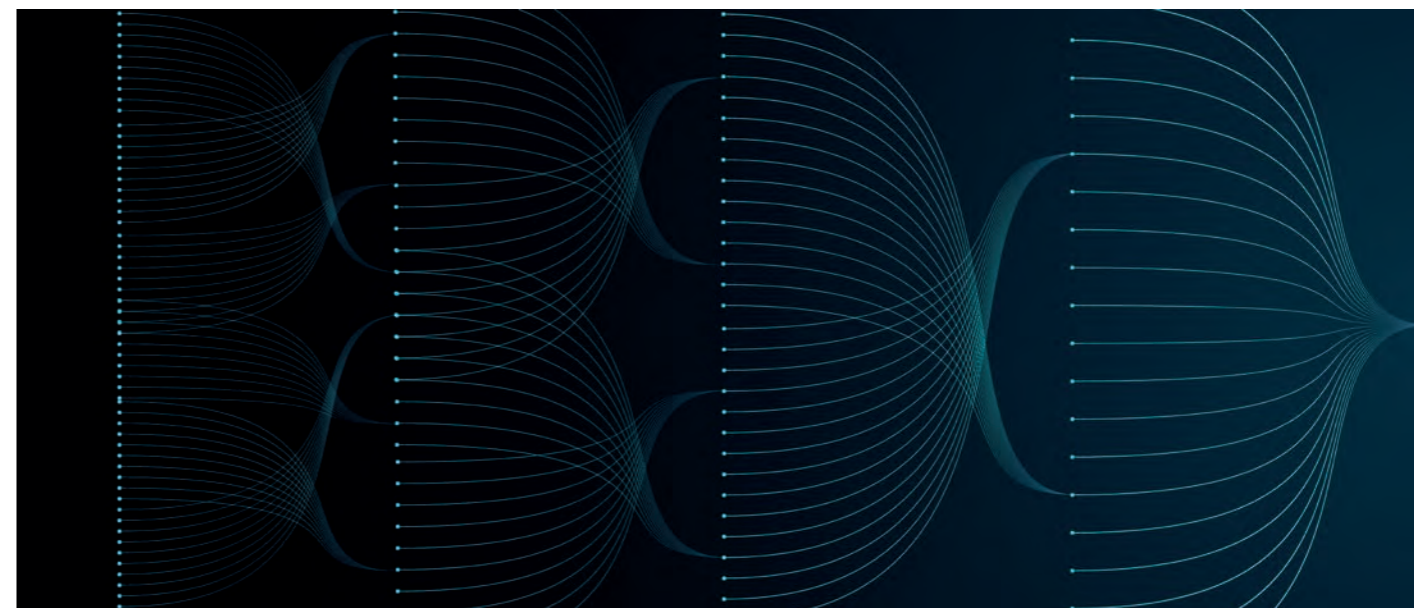


Artificial intelligence technologies

Trustworthy, safe AI

Verification of neural networks: a challenge to overcome

The formal verification of neural networks presents many challenges. Although there are languages to describe how a neural network is expected to behave, current validation tools do not consider the full richness of these languages. CEA-List is investigating practices from the field of programming languages to expand the scope of what can be formally verified.



Critical software systems are essential to our society. They help ensure the safety of everything from railway networks to nuclear power plants. Software malfunctions can have grave consequences for our infrastructures, the environment, and individuals. Software can only be implemented in these critical environments if there is a high degree of trust that it will run as intended. The study of programming languages—a branch of computer science—focuses on developing new ways of designing and coding software to make it safer.

Neural networks are now making inroads into critical software, and the rules of the game have changed. The way neural networks are designed makes them difficult to describe formally and, therefore, to verify. Python, the most widely used programming language for neural network design, is very difficult to analyze. To further complicate matters, neural networks are made up of series of very large algebraic calculations with no obvious link to the result produced. What this means is that, in terms of their characteristics, neural networks are very different from the conventional software the verification community is accustomed to working with.

Our research began with a description of the effects of neural network verification on program verification tools. We made two crucial observations:

- First, the test benches currently used to verify neural networks are biased toward a very specific class of problems, which appear to us to be far removed from the needs of industrial end-users.
- Second, these verification tools work with very-low-level inputs (close to machine language). This creates a very wide and unaddressed gap between the formal description of a program and what it is possible to verify (Figure 1).

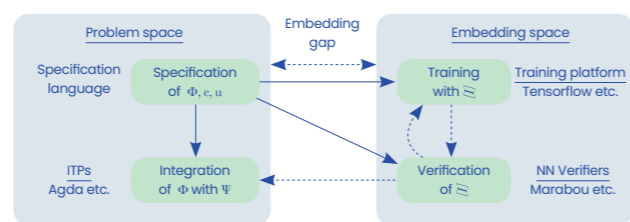


Figure 1: The distance between a program's formal specification and its actual behavior.

Based on these observations, we decided to adapt practices already known in programming language research to neural networks:

- Describing the expected safety and security; properties directly in the programming language and;
- A more realistic approach of looking at the literature on specification languages to formally describe existing systems. Part of this formal description, more readily usable by existing formal verification tools, was added to the program itself.

This enabled us to propose description languages suitable for this approach.

« The specification of neural networks is an important research topic that must be addressed at the fundamental level of the programming language.»

••• JULIEN GIRARD-SATABIN

••➤ **Application**
CAISAR software platform

••➤ **Major projects/partnerships**
SAIF project (PEPR IA), DeepGreen project.

••➤ **Flagship publication**
«Neural Network Verification is a Programming Language Challenge». L. C. Corneiro, M. L. Daggitt, J. Girard-Satabin, O. Isac, T. T. Johnson, G. Katz, E. Komendantskaya, A. Lemesle, E. Manino, A. Sinkarovs, H. Wu, *In European Symposium on Programming (ESOP), 2025.*
<https://doi.org/10.48550/arxiv.2501.05867>



Artificial intelligence technologies

Trustworthy, safe AI

Uncertainty in AI-guided Monte Carlo simulations

AI models can speed up the simulation of materials at the atomistic scale. However, the models' approximation introduces uncertainty that biases the results. We developed a method, PEM, to adapt the Monte Carlo algorithm and penalize regions that are uncertain for the AI, mitigating AI errors and making deep-learning-based materials modeling trustworthy.



EIJI KAWASAKI
Research Engineer and Expert



DIMITRIOS TZIVRAILLIS
PhD student

In our research on complex materials, we used multi-scale approaches that establish a relationship between a material's atomistic behavior and macro-scale properties. The goal is to bridge the gap between the atomistic scale, which is governed by interatomic forces, and the macroscale, with a view to modeling materials and their defects accurately.

Molecular dynamics, Monte Carlo sampling, and other simulation methods remain essential. However, because they require repeated and precise total energy and force calculations, they quickly reach their limits for the modeling of large or complex systems. Researchers use deep learning models—machine learning potentials (MLP)—as surrogate functions to get around the high computational cost of the simulation methods mentioned above.

These neural networks effectively replace high-fidelity physical calculations by efficiently approximating potential energy surfaces or force fields. This significantly speeds up atomistic simulations, creating new possibilities for the exploration of a wider range of configurations. However, the increase in speed creates a new challenge around trustworthy AI. Because they are approximations, MLPs inherently come with errors like epistemic uncertainty, which increases when the model is faced with configurations that have been encountered little or not at all during training. Our research revealed that this energy

uncertainty can distort predictions of macroscopic physical properties. The most important finding from this research is that, if this uncertainty is to be recognized and actively managed, adaptations must be made to the simulation algorithm itself.

We developed the Penalty Ensemble Method (PEM) to quantify this uncertainty and integrate it directly into the Monte Carlo simulation. The standard Metropolis-Hastings acceptance rule that MC-MC is based on determines whether a proposed configuration is accepted or rejected. PEM factors in uncertainty to modify the rule. This increases the probability of rejection for configurations located in regions where MLP exhibits high predictive uncertainty.

By penalizing displacements toward uncertain regions, PEM mitigates the bias introduced by the MLP error, making the simulations more reliable and consistent. Despite the use of an approximate model, the properties obtained are close to the reference values.

We demonstrated that using PEM to integrate uncertainty quantification provides a solid path toward combining the speed advantages of deep learning with the physical precision of Monte Carlo simulations. This advance paves the way toward robust, reliable materials modeling to meet the needs of critical industries.

Flagship publication
«The PEM method» was developed in 2025 as part of Dimitrios Tzivraillis' PhD research with Alberto Rosso, Research Director at the Laboratoire de Physique Théorique et Modèles Statistiques (LPTMS, CNRS).
<https://doi.org/10.48550/arXiv.2506.14594>

Major project
Tzivraillis' PhD research is funded by the PRIMA (Probabilistic Reasoning in Machine Learning) project to extend the operational domain of AI by providing usability guarantees for deep learning models. One of PRIMA's primary objectives is the quantification of uncertainty: calculating and controlling uncertainties (both random and epistemic) to make AI a reliable scientific calculation tool.

Application
PRIMA is part of our research on trustworthy AI, essential for critical use cases like materials modeling and simulation-based inference (SBI).

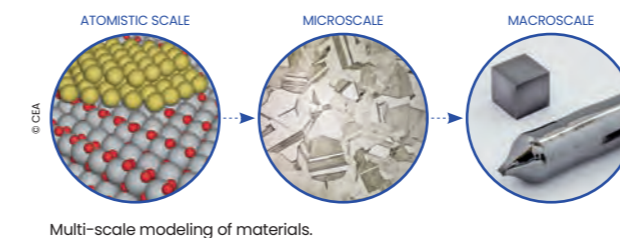


Fig. 1: Microscale energy predicted as a function of a precise calculation of energy. The AI's excellent prediction performance is worth noting.

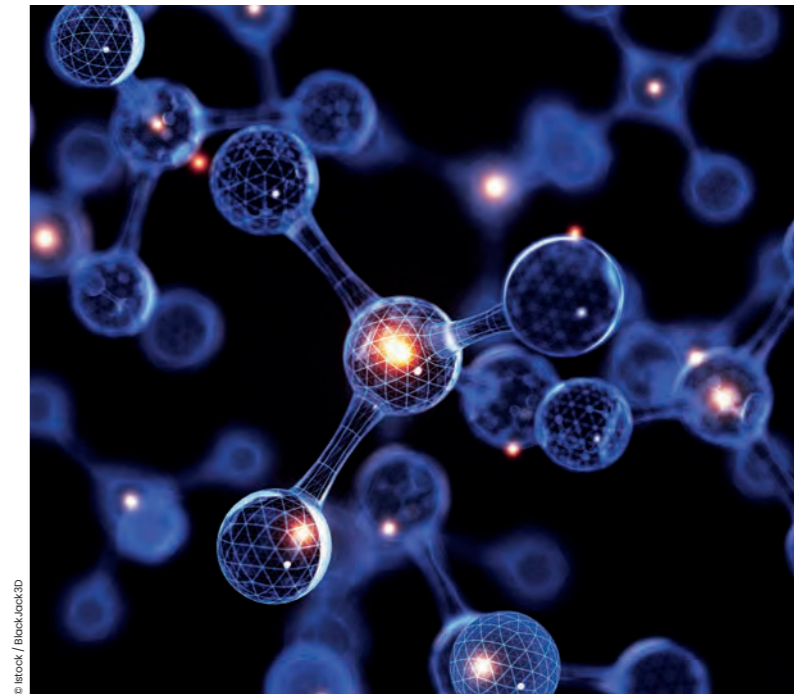
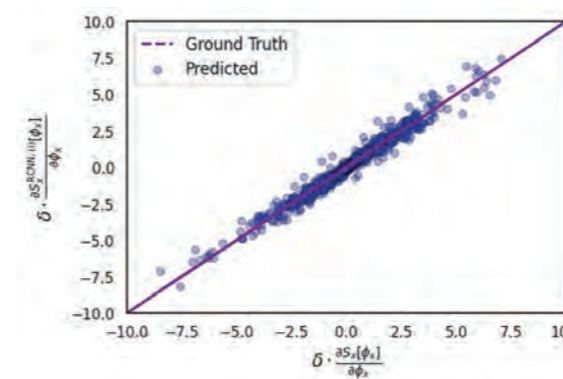
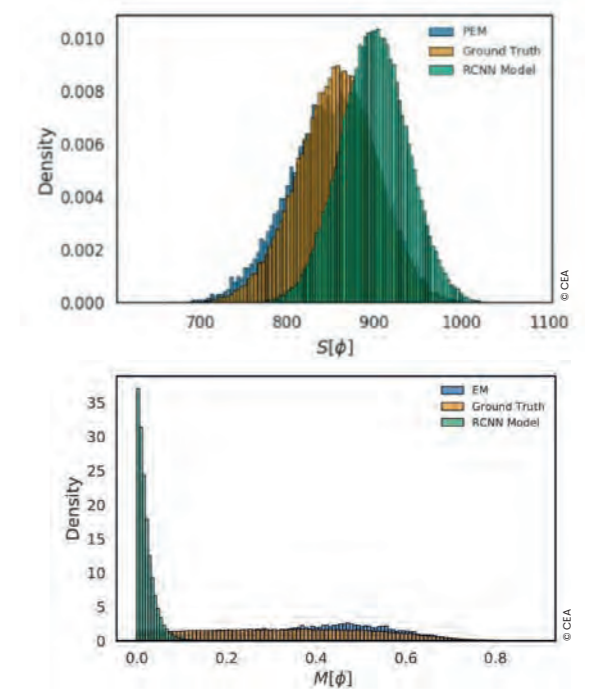
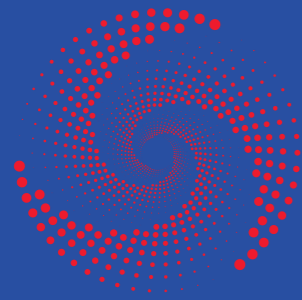


Fig. 2: Monte Carlo simulation results. Despite the AI models' excellent performance, significant bias between precise and costly simulation (orange - Ground Truth) and AI-guided simulation (green - RCNN) was observed. The AI-guided simulation can be corrected using the PEM method (blue).



03



Computing and distributed systems

When it comes to computing on large volumes of data in a way that is safe, trustworthy, and environmentally responsible, distributed systems offer several solutions. Our distributed systems research reflects our holistic vision of a continuum encompassing hardware, software, theoretical models, and design tools.

Computing and distributed systems present new challenges around managing increasingly large volumes of data while delivering extremely high computing performance and guaranteeing safety, trustworthiness, and environmental responsibility. We are working toward a variety of scientific objectives to overcome these challenges. For example, we are designing auditable systems and rigorous models to ensure that these systems are reliable and secure while safeguarding against the leakage of sensitive data. Innovative digital architectures—like 3.5D architectures and advanced architectures based on emerging memories—are another area we are exploring to break through performance and power bottlenecks.

Finally, we are integrating new paradigms like quantum computing as a promising path toward solving complex optimization problems. Our vision for the future is one in which distributed architectures and algorithms form a coherent continuum spanning hardware, software, theoretical models, and design tools. The research highlighted in this chapter illustrates how we are making this vision a reality, tackling challenges from different angles that include distributed systems theory and engineering, advanced architectures based on emerging memories and 3.5D integration, attack simulation and analysis, and quantum algorithms.

52 Distributed systems

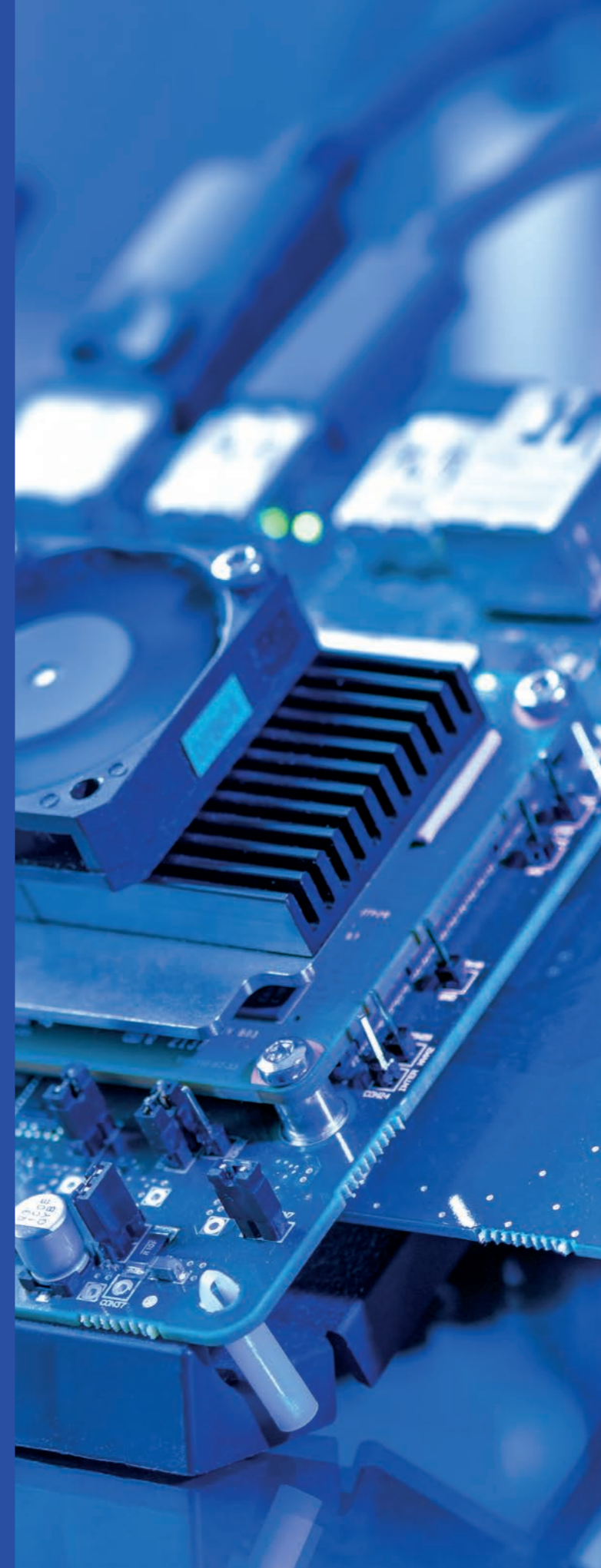
- 52 Auditing data access with no leaks
- 54 Simulation of attacks on distributed systems

56 Hardware architectures

- 56 Making 3D integration a reality
- 58 Embedded non-volatile memory with robust, optimized accesses

60 Quantum

- 60 Quantum computing applications and use cases
- 62 Qbricks: toward the formal verification of hybrid classical/quantum programs



Computing and distributed systems

Distributed systems

Auditing data access with no leaks



The ability to provide a complete data access audit is vital for critical systems that handle sensitive data. CEA-List contributed to the development of a new formal framework that guarantees that every read access operation—even if interrupted—is detected while protecting the access logs from prying eyes.

In distributed systems, several remote computers cooperate to perform tasks. One of the major security requirements for systems that handle sensitive data (healthcare, finance, legal...) is the ability to keep reliable data access logs. This presents two main challenges: The first is to guarantee that every access operation is detected, ensuring that no users go unaudited; and the second is to do this in a way that protects the access logs from prying eyes.

In the field of distributed systems, the first formal (i.e. mathematically rigorous) definitions of auditability are fairly recent. And they are based on the strong assumption that a read access operation is only audited if it is completed. This is a loophole that bad actors could use to access data and then simulate a malfunction that prevents completion, thus avoiding detection. Plus, if the access logs are available to anyone, bad actors could use this information to figure out who accessed which data—creating problematic data leaks.

In a departure from the existing definition of auditability for distributed systems, CEA-List introduced two new formal definitions used as a conceptual framework for this research. The first definition characterizes what constitutes an effective read access operation as the instant a reader might learn the value being accessed, regardless of whether the read access operation is completed or not. The second definition establishes the concept of an

“uncompromised operation”, which guarantees that it cannot be known if an operation was completed on the data or not.

From an algorithmic standpoint, access to a piece of data and the recording of the access operation in the audit log are combined in a single, indivisible operation, ensuring that a read access operation is logged, even if it is not completed. To prevent readers from learning about each other's accesses, each reader is associated with a bit (0 or 1) initialized secretly and at random, creating a simple but unbreakable encryption to protect the audit log. When a read operation is performed, the reader's bit is inverted (0 becomes 1, or vice versa). The values observed look like a random sequence of 0s and 1s, making it impossible to determine which readers have accessed the data. Only auditors who have the initial secret key can know who has read which data. This is done by comparing the key with the final log.

Formal proof of the algorithm guarantees reliable operation of the system in all circumstances. Despite multiple access operations, the data remains consistent (when a piece of data is read, the last value assigned to it is clearly visible). In addition, bad actors cannot interfere with the system, all read operations are effectively logged, and the privacy of the access operations is protected, even if curious readers look at the audit logs.



“ Our solution provides full and private traceability every time data on a distributed system is accessed.”

••• ALEXANDRE RAPETTI

••• **Flagship publication**



«Auditing without Leaks Despite Curiosity».

ATTIYA Hagit, FERNANDEZ ANTA Antonio, MILANI Alessia, RAPETTI, Alexandre, & TRAVERS, Corentin. In *Proceedings of the ACM Symposium on Principles of Distributed Computing. 2025. p. 455-465.*

<https://doi.org/10.48550/arXiv.2505.00665>

This research was published in the proceedings of PODC, the leading international conference (A-rank*) on distributed algorithms.

Figure 1: Data access scenarios

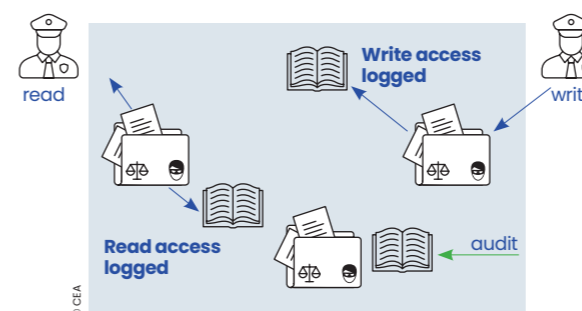
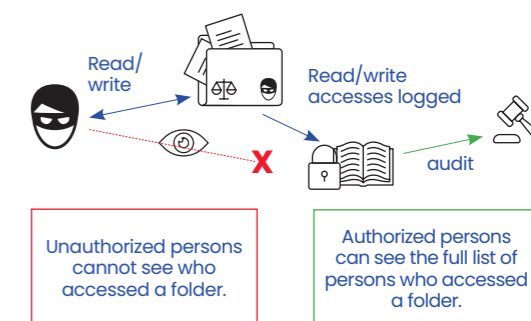


Figure 2: Secure audit



Computing and distributed systems

Distributed systems

Simulation of attacks on distributed systems

How do you test the robustness of a distributed system against realistic attackers? To round out the more conventional theoretical analyses used to assess robustness, we integrated detailed simulations into our multi-agent tool, MAX, to determine the real impact of attacks on security.



“ With MAX, we have been able to simulate and measure the success of attacks on various distributed systems.”

•• ERWAN MAHE

“ One of EDF R&D’s scientific objectives is to simulate and test the resilience of large distributed systems in conditions as close to real operations as possible. This research helped us achieve this objective.”

•• PIERRE-YVES PIRIOU

The control of much of the internet by a small number of companies has raised increasing societal, economic, and geopolitical concerns—and some very valid questions about cybersecurity. Some of these issues could be addressed by new Distributed Ledger Technologies (DLT), members of the blockchain family. Unlike centralized, hierarchical systems, DLTs are secure systems where information is replicated on every machine in the network and decisions, made by a subset of users on the network, are distributed. This setup makes it more difficult for attackers to take control of the system.

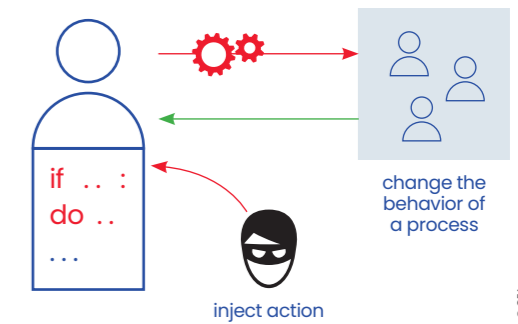
Despite these advantages, the adoption of DLTs has been hindered by the high power consumption and hardware requirements of replicating data and communicating between machines on the network.

New algorithms that attempt to address these issues are developed regularly. However, because of the complexity of these systems and the properties that need to be verified, security verification remains a challenge.

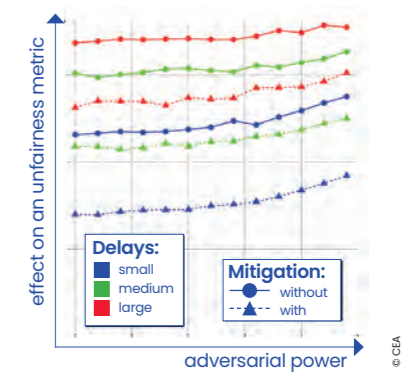
CEA-List developed its MAX software platform to simulate and assess DLT behavior in a variety of conditions. For example, the underlying communication network can be configured to assign a probabilistic distribution of the routing times of messages between two actors. Simulations of bad actors doing things like creating or intercepting messages or taking control of other machines on the network can also be configured. MAX can be used to assess properties like the consistency of local data between each machine. Metrics for assessing system behavior—like the order in which queries were submitted or processed—can also be extracted from the simulations.

These metrics can then be used to assess the relative robustness of different decentralized systems under different attacks and test how effective different countermeasures are against these attacks.

MAX can be used to assess properties like the consistency of local data between each machine. A variety of metrics for quantitatively assessing system behavior can also be extracted from the simulations. One such metric consists of measuring the distance between the order in which requests were submitted to the system and the order in which they were processed, for example.



An adversary takes control of an actor



MAX simulations: CEA-List countermeasure (triangles) improves digital ledger technology (DLT) resilience

•• Contributors
Erwan Mahe, Rouwaida Abdallah, Sara Tucci-Piergiorganni.

•• Projects
Our most recent research using MAX was conducted for the CyberQL project with EDF.

•• Flagship publications
This research resulted in two publications:
“Adversary-Augmented Simulation to evaluate order-fairness on HyperLedger Fabric” <https://doi.org/10.48550/arXiv.2403.14342> at LADC 2024.

“Order Fairness Evaluation of DAG-based ledgers” <https://doi.org/10.48550/arXiv.2502.17270> at BCCA 2025. We also presented a MAX quick-start tutorial at BCCA 2025.

Computing and distributed systems

Hardware architectures

Making 3D integration a reality

The CEA pioneered 2.5D and 3D integration with its IntAct circuit^[1] and has held on to its position as a leader in 3D integration ever since. Recent research by two CEA-List PhD students has led to significant advances in communications and security. These two solutions, published in a special issue of the IEEE Journal of Electronic Testing: Theory and Applications (JETCAS), are a testament to the CEA's commitment to innovation and the development of competitive technologies.



2. Enhancing DFT Security in Chiplet-Based Systems with Encryption and Integrity Checking

Multi-chiplet systems come out of complex and fragmented ecosystems involving chiplets manufactured by different foundries and then tested and assembled by other third-party companies. Despite this complexity, each chiplet must be authenticated to prevent counterfeit or malicious chiplets from being integrated into systems and to ensure that sensitive data cannot be exfiltrated through test ports. PhD student Juan Suzano's research, financed by STMicroelectronics and conducted in partnership with STMicroelectronics and TIMA (a Grenoble INP-Grenoble Alpes University-CNRS joint research unit), resulted in an innovative solution combining chiplet authentication and test data encryption. The solution developed, compatible with IEEE standards 1686 and 1838 and designed for industrial deployment, was published in JETCAS^[3]. This concrete response to the vulnerabilities of chiplet-based architectures will support the development of an open, secure ecosystem.

« 3D integration is now a pillar of high-performance system architectures.»

•• DAVY MILLION

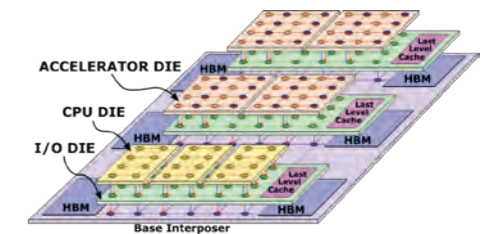


Figure 1: Example of a system with 3.5D topology

TWO ARTICLES ON 3D INTEGRATION IN IEEE JETCAS

1. Depth-First: A Deterministic and Scalable NoC Routing Protocol for 3.5D Packaged Architectures

3.5D (also referred to as 5.5D) integration is revolutionizing circuit architectures: Instead of expanding chips horizontally, multiple chiplets are stacked on an interposer, in vertical miniature skyscrapers. Networks-on-chip (NoCs) with special routing algorithms allow flits (small units of data) to travel from a source to a destination, ensuring communication between the chiplets (processors, GPUs, accelerators...) in the stack. The routing algorithms used in 2.5D systems (in which several chiplets are stacked directly on an interposer) are well established. However, in 3.5D architectures, flits can take several vertical routes between chiplets to reach their destinations, introducing additional complexity that requires innovative solutions. Fulbright Scholar Davy Million developed the **Depth-First**^[2] algorithm specifically for 3.5D architectures as part of his PhD research in conjunction with the University of California at Santa Barbara. The algorithm routes data efficiently, avoiding deadlocks (circular dependencies between network resources that can cause the network to stop functioning), marking a major advance for 3.5D circuits.

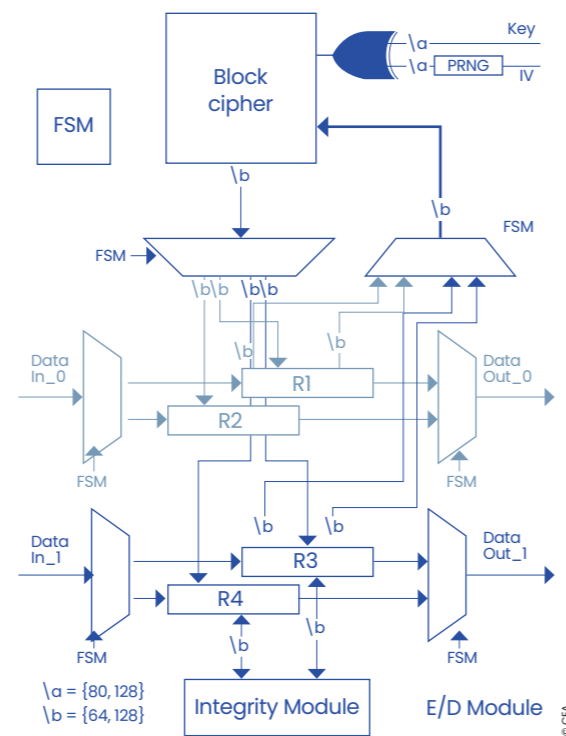
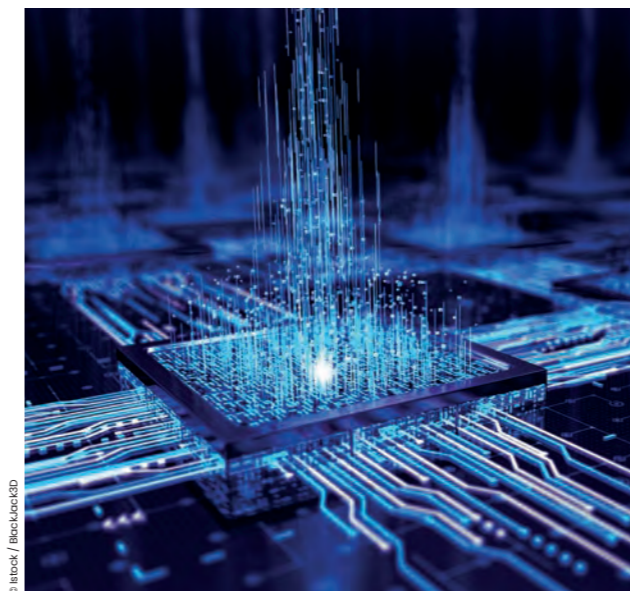


Figure 2: Secure test architecture compatible with IEEE 1838 and 1687

The market for chiplets and 3D circuits is expanding at a 24% CAGR and is expected to reach \$157 billion by 2030.

24%



•• Applications
2.5D, 3D, 3.5D circuits.

•• Patents
Two chiplet authentication patents pending.

•• 3.5D NoC simulator on CEA-List's GitHub:
<https://github.com/CEA-LIST/Noxim-3.5D>



•• Flagship publications

[1] «IntAct: A 96-Core Processor With Six Chiplets 3D-Stacked on an Active Interposer With Distributed Interconnects and Integrated Power Management», P. Vivet et al., in *IEEE JSSC*, vol. 56, no. 1, pp. 79-97, Jan. 2021.
<https://doi.org/10.1109/JSSC.2020.3036341>

[2] «Depth-First: A Deterministic and Scalable NoC Routing Protocol for 3.5D Packaged Architectures», D. Million et al., *IEEE JETCAS* 2025.
<https://doi.org/10.1109/JETCAS.2025.3590106>

[3] «Enhancing DFT Security in Chiplet-Based Systems With Encryption and Integrity Checking», J. Suzano, et al., in *IEEE JETCAS* 2025.
<https://doi.org/10.1109/JETCAS.2025.3592984>



Computing and distributed systems

Hardware architectures

Embedded non-volatile memory with robust, optimized accesses

The CEA has been conducting R&D on resistive memory (ReRAM)—an alternative to flash memory—for more than a decade. A complete system integrating ReRAM into a processor architecture was recently designed and validated for advanced memory company Weebit Nano (Figure 1). Efficiency and energy-delay product gains of 100x and 15x, respectively, were obtained over one million cycles through algorithmic improvements to the calibration of read and write access operations. This research lays the foundation for the next generation of memories Weebit Nano is currently commercializing under license to several foundries.

Israel-based Weebit Nano, founded in 2015, is addressing the memory market with innovative solutions. The company turned to the CEA to develop a complete demonstrator to promote its patented eNVM (embedded non-volatile memory) technology. Weebit Nano has already used the demonstrator to sell or license its technology to foundries in the US and Asia. The CEA's knowledge of eNVM, circuit and systems design, and wafer- and package-level testing was what originally motivated Weebit Nano to explore the potential for a partnership.

The demonstrator produced includes a non-volatile **ReRAM macro** and analog circuits that provide the precise, yet tunable signals required for this type of memory's complex read and write operations (Figure 1). It also includes control circuits based on RISC-V subsystems, as well as a dedicated controller developed specifically for this project.

The Smart Write Algorithm (SWA) describes how the solution works (Figure 2). Various write-assistance techniques were cleverly implemented to improve the intrinsic characteristics of the ReRAM fabricated at the

CEA. To achieve this, the system built had to be flexible enough to regulate a large number of current and voltage signals in the ReRAM macro, yet observable enough for detailed characterization. The design techniques used include (Figure 1):

- **Analog techniques** limiting or detecting when a threshold current is reached during write operations.
- **Digital techniques**, such as:
 - Reading the binary value of a memory cell before each write operation and writing only if the bit is not in the desired state.
 - Checking the state of the cell after the write operation and repeating the write operation until it is correct.
 - An error correction code (ECC) that detects and/or corrects a word read using bits encoding the parity of the word's subgroups.

This research and, especially, the optimizations obtained by using the Smart Write Algorithm, SWA, led to significant improvements (Figures 2 and main) at room temperature:

- **Functionality:** 1 million error-free read and write cycles, 10x better than the current state of the art in ReRAM in the literature.



BASTIEN GIRAUD
Research Engineer



ILAN SEVER
Vice President Research & Development
WEEBIT NANO LTD.

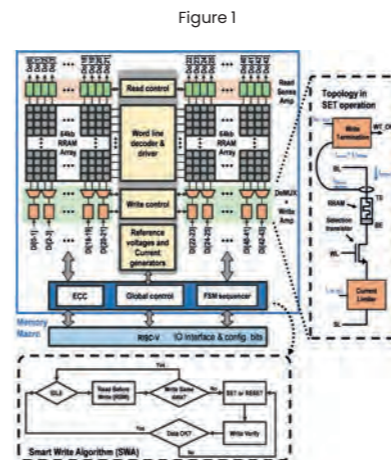


Figure 1

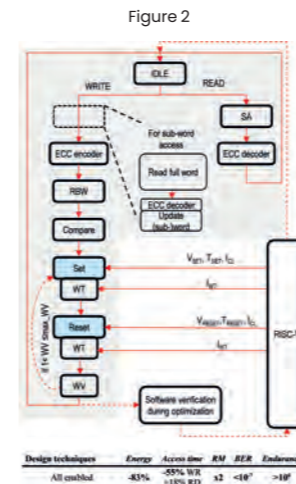


Figure 2

- **Performance:** 83% lower energy consumption and 55% faster access during the write phase compared to implementations with no write-assistance techniques. This corresponds to an 8x reduction in the energy-delay product with ECC (and 15x without).
- **Yield:** 100x lower bit error rate (BER) and a 2x read window, positioning the solution well beyond the state of the art both at room temperature and at high temperatures (85 °C).

These improvements make the ReRAM technology developed with Weebit Nano a promising candidate for widespread industrial deployment.

15x improvement in energy-delay product.

15X

« This exciting project, which produced the CEA's first NVM design, involved seven CEA laboratories from technology to circuit design to system-on-chip testing.»

• • • **BASTIEN GIRAUD**

« We are extremely happy with the team's work... Meeting the schedule for the tapeout, adapting to Weebit change requests, implementing multiple test-modes and flexible design, extremely clean layout.»

• • • **ILAN SEVER**

- • • **Use cases, applications, technology transfer**
The ReRAM IP, licensed to the customer, can now be transferred to the customer's partner foundries (SkyWater, OnSemi, DB HiTek...).

- • • **Patents**
Weebit Nano holds licenses for five patents for the Smart Write Algorithm (SWA), the subject of two publications (JSSC and IMW), and for the error correction codes and signal sequencing during read operations.

- • • **Major project and/or partnership**
This 130 nm design project strengthened the reputation Weebit Nano has built up over many years. It also led to several projects on ECC, security analysis, 22 nm system design, and, coming soon, in-memory computing for AI.

- • • **Flagship publications**

 «Smart Write Algorithm to Enhance Performances and Reliability of an RRAM Macro», B. Giraud and al., in IEEE Journal of Solid-State Circuits, vol. 59, no. 9. <https://doi.org/10.1109/JSSC.2024.3386429>
 «Benefits of design assist techniques on performances and reliability of a RRAM macro», B. Giraud and al., IEEE IMW, 2023.

Computing and distributed systems

Quantum

Quantum computing applications and use cases

Much like classical application-specific processors, quantum processing units (QPUs) are expected to be used to speed up certain computational steps in algorithm execution.

CEA-List is focusing on two main issues:

The first is to determine the applications QPUs are most efficient at handling; and the second is to benchmark the performance of different QPUs on different applications.

For QPUs to speed up computing, the problems of interest must first be converted into a series of instructions, or queries, that can be processed by the QPU. For many applications, formatting problems as queries is a bottleneck that can reduce and, in some cases, cancel out any increase in speed. Our laboratory is investigating strategies to prepare and use data for computing on QPUs that would overcome this bottleneck. Our research recently led to two publications:

1. For gate-based QPUs: The query takes the form of a matrix (which must be expressed as a product of the smallest possible number of elementary quantum gates) of the problem's data. Here, we focused on finding the most efficient way to break down application-specific matrices. One of the articles presented this year^[1] at IPDPS proposed improvements to the implementations of a number of matrices. We also developed code to automate the techniques described in the article. A paper detailing this development has been submitted to APP.
2. For continuous-time QPUs (also referred to as analog QPUs or quantum simulators): These machines can handle any problem that can be reduced to a quadratic unconstrained binary optimization (QUBO) problem. Here, we proposed effective reductions for application-specific problems. This led to a new reduction for solving linear systems, which in turn led to a patent.



STÉPHANE LOUISE
Research Director



FRÉDÉRIC BARBARESCO
Leader, Quantum Algorithms and Computing and AI and Algorithms for Sensors segments
THALES TRT

We are also investigating quantum optimization heuristics and how to benchmark them. Unlike benchmarks based on the quality of quantum operations, here, we are comparing the performance of quantum processors with that of other types of processors—a particularly relevant issue. We initially focused our benchmark on two technologies, chosen for their maturity, D-Wave's quantum annealers (QA) and NEC's quantum-inspired vector annealing (VA) heuristic. We were thus able to compare the performance of quantum heuristics with each other and with classical heuristics. This benchmark, the first to compare the heuristic performance of QA and VA on a constrained problem, was presented^[2] at QCE.

This research is one of the CEA's contributions to the BACQ project, part of the MetriQs-France program financed through the national France 2030 initiative. The approach developed will soon be applied to Pasqal's Ruby and Quandela's Lucy quantum computers at the CEA's high-performance computing center. The new knowledge of quantum annealers led to contributions to summer (EDF-CEA-Inria) and winter (Riken) schools and to tutorials (Hub quantique).

- • > **Use cases**
Combinatorial optimization and linear system resolution (simulator-type machine) and chemistry, n-SAT, PDE and hollow matrix (gate-based machine).
- • > **Patent**
Patent pending, publication submitted to APP for future OSS (SCB-generator).
- • > **Partnerships**
LNE, Thales, Eviden.
FZ Jülich (AIDAS project), Riken.

• • > **Flagship publications**

[1] «Gate Efficient Composition of Hamiltonian Simulation and Block-Encoding with its Application on HUBO, Chemistry and Finite Difference Method», R. Ollive and S. Louise, 2025 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPS), Milano, Italy, 2025, pp. 519-528, <https://doi.org/10.1109/IPDPSW66978.2025.00083>

[2] «Vector Annealing, a Quantum-Inspired Technique: Benchmarking Performance Against Quantum and Simulated Annealing within the BACQ Framework», S. Louise, 2025 IEEE International Conference on Quantum Computing and Engineering (QCE), <https://doi.org/10.1109/QCE65121.2025.00217> the leading conference in the field.

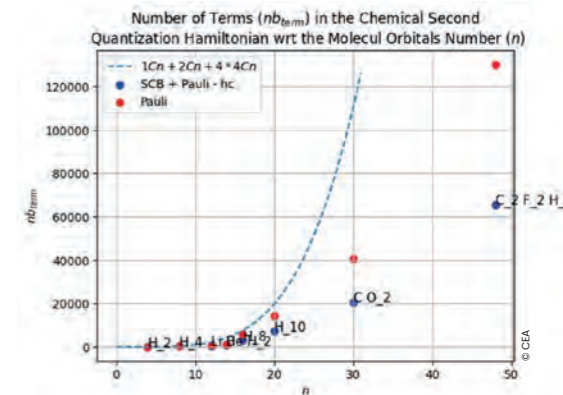


Figure 1: Improvements obtained for encoding quantum chemistry problems on a gate-based architecture (Q-LCYL: blue dot, usual: red dot).

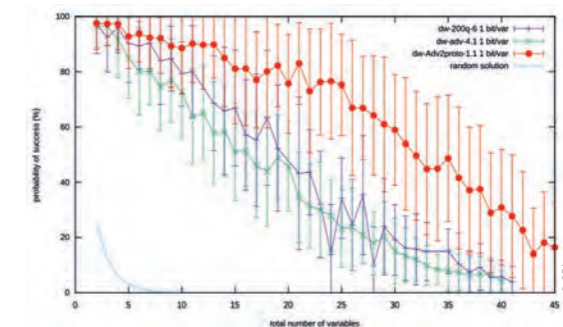
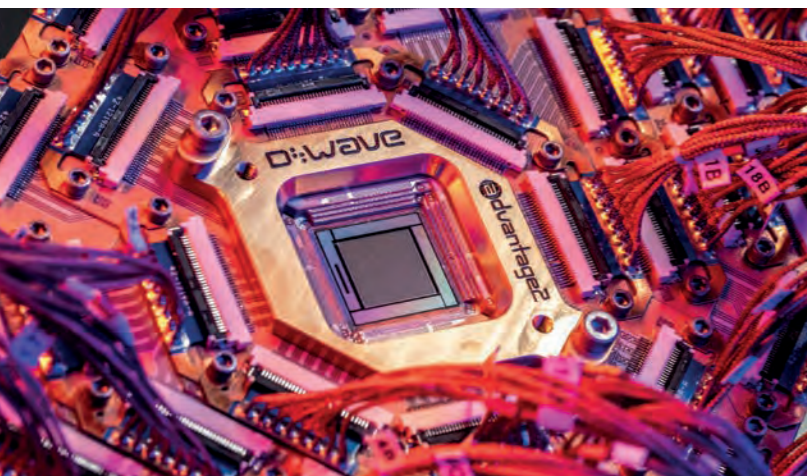


Figure 2: Brute-force linear-system computing capacity using Quadratic Unconstrained Binary Optimization (QUBO) as a function of the number of variables on different generations of D-Wave QPUs



Advantage 2 quantum processing unit.

« Although quantum computers (QPUs) are still at a low TRL, early benchmarks show that significant progress is being made from one generation of QPU to the next »

• • • STÉPHANE LOUISE

« BACQ for Quantum Computer Benchmarking: If you cannot measure it, you cannot improve it (Lord Kelvin) »

• • • FRÉDÉRIC BARBARESCO

Computing and distributed systems

Quantum

Qbricks: toward the formal verification of hybrid classical/quantum programs

Advances in quantum computing are creating new challenges to our current code validation practices. CEA-List is developing new formal-analysis-based verification techniques to respond to this new context. Most of the programs run on tomorrow’s quantum computers will be hybrid, and our recent research focuses on these programs, which combine classical and quantum computing.



CHRISTOPHE CHARETON
Research Engineer



MATHIEU NGUYEN
Research Engineer



JAD ISSA
PhD student

Quantum program validation

The main principle behind software testing is to verify the runtimes of the expected system properties. These program validation and verification techniques cannot simply be transposed to quantum programs, however. Because of the physics at work inside quantum computers, the value of a unit of data cannot be read at runtime without destroying it. The non-deterministic nature of quantum computing also makes scaling up test strategies impossible. Therefore, for any future quantum programming scenario to be realistic, new and innovative code verification methods will have to be developed.

Over the past several years, CEA-List has been developing a formal-proof-based quantum program verification tool called Qbricks. In Qbricks, a quantum program’s expected specifications can be written explicitly. The tool’s automated and interactive proof features provide verification that the program behaves according to specifications on all possible input data.

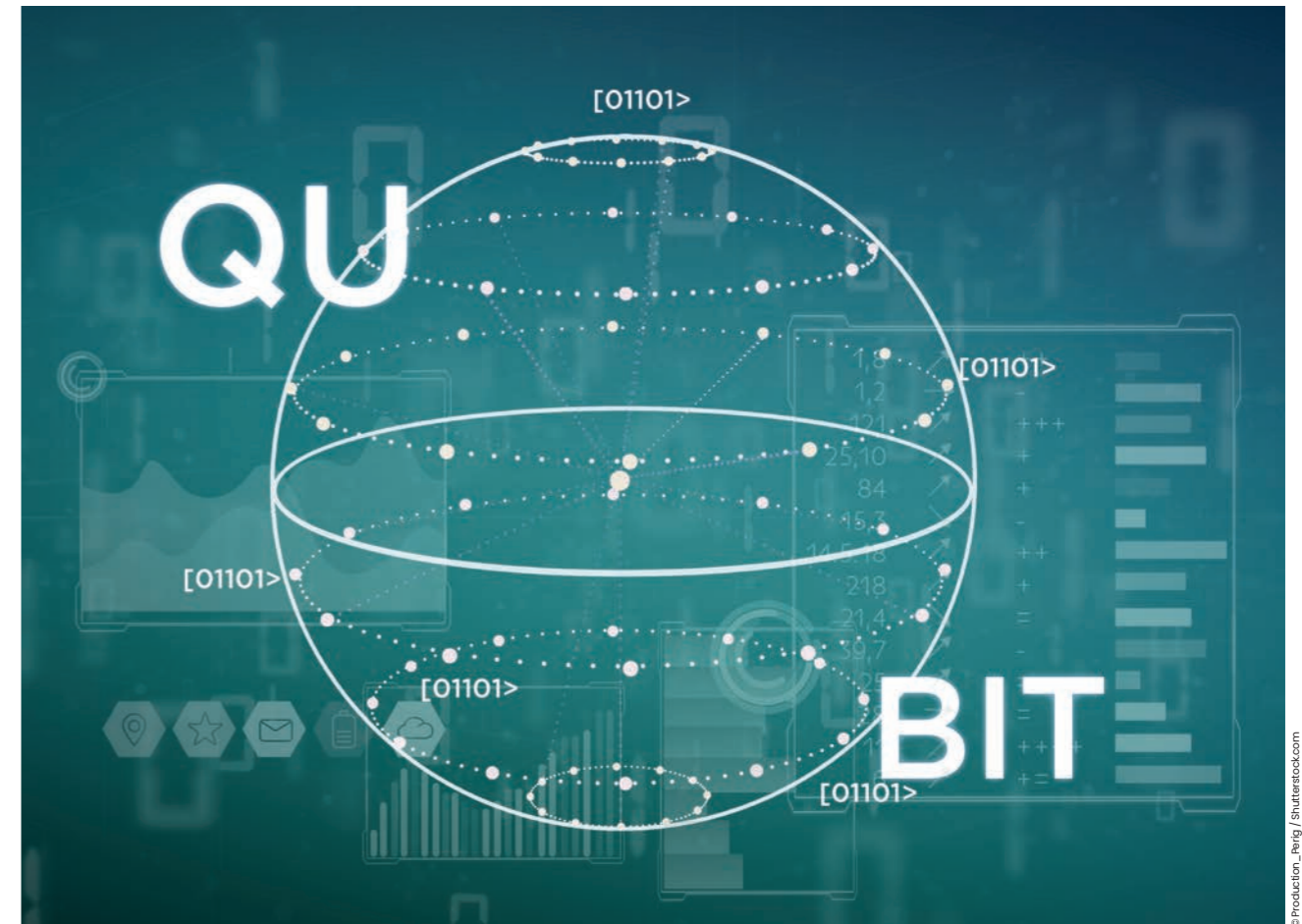
Extending the analysis to hybrid programming

Our most recent research focuses on extending this principle to hybrid programs, in which classical and quantum data interact.

There is abundant research on hybrid computing in the literature. The quantum variational eigensolver (QVE) algorithm and other algorithms that combine classical and quantum computing components or that embed quantum computing in classical structures have been amply studied. In addition, in the general case, the quantum phase estimation (QPE) algorithm, Shor’s quantum factoring algorithm, and most other purely quantum algorithms perform better when broken down into simpler components and hybridized. In large-scale quantum computing, even purely quantum primitives are implemented in a hybrid manner—a prerequisite for error detection and correction.

We started by extending equivalency verification—crucial to validating program compilation and optimization—from purely quantum computing to different classes of hybrid computing.

To address the general case, however, the symbolic program representation system used in Qbricks will need to be overhauled. This research is ongoing and encouraging results have been obtained on symbolic circuit execution experiments.



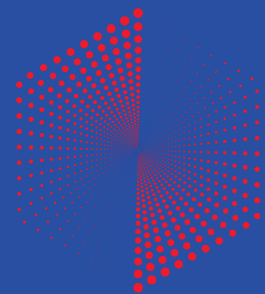
“ **New code verification and debugging methods will be required for any realistic future quantum programming scenario** ”

••➤ **CHRISTOPHE CHARETON**

••➤ **Projects**

This work is part of the Quantum PEPR and the HQI initiative in partnership with Paris-Saclay University, Loria, and Grenoble Alpes University. DGA (France’s military procurement agency), Eviden, and Thales have also shown a marked interest.

04



Digital instrumentation and the factory of the future

Digital instrumentation and robotics are becoming increasingly common in industrial systems. This changing landscape is creating some major new scientific and technological challenges, not least of which is how to ensure the reliable acquisition and use of data in these complex environments.

The digitalization of production systems depends heavily on digital instrumentation, which will require the development of robust, smart, communicating sensors and their integration into demanding environments. Add to this the need to process and use data in real time, and you get a whole set of new problems to solve. Robotics presents some additional challenges around multi-sensor data fusion, environmental perception, and safe, efficient interaction and cooperation between humans and industrial systems.

CEA-List's research directly addresses this changing industrial environment. We are developing advanced instrumentation, measurement, and AI technologies and methods specifically for advanced systems. Our research topics range from cognitive robotics and planning for robotic gripping to innovative measurement techniques for severe environments in manufacturing, nuclear power, healthcare, and science.

We are helping make industrial systems more efficient, flexible, and resilient through new approaches to the physical measurement, acquisition, processing, and interpretation of complex data and the integration of data into decision-making and control systems. Our goal is to develop tools that are safe and easy to learn and understand, keeping humans at the center of supervision and decision-making processes.



66 Robotics and interfaces

- • 66 Estimation of a multifingered gripper joint configuration using 3D point clouds and AI
- • 68 A cognitive programming interface
- • 70 Active lubrication, a kinesthetic and vibrotactile haptic brake

72 Sensors

- • 72 Fiber Bragg gratings for ultrasonic measurement at high temperatures
- • 74 Quantum sensors lift the veil on neutrinos
- • 76 Electrochemical measurement techniques for nitrite detection

78 Modeling and characterization

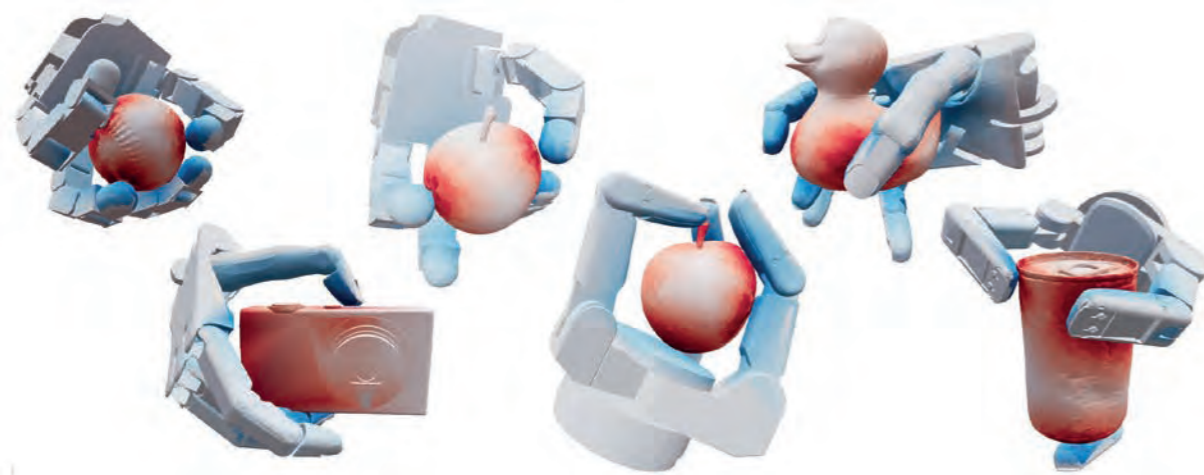
- • 78 Continuous eddy-current monitoring of fatigue degradation in an unnotched steel sample
- • 80 AI algorithms for gamma-ray spectrometry used in field measurement
- • 82 Photoneutrons and photofission, scientific weapons against trafficking

Digital instrumentation and the factory of the future

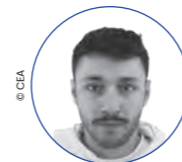
Robotics and interfaces

Estimation of a multifingered gripper joint configuration using 3D point clouds and AI

In robotic gripping, the positions of the effectors (phalanges or fingertips) on an object must be controlled precisely. Knowledge of the joints' internal configurations is essential to achieving this. We looked at a practical case where the only information available was a 3D point cloud of the gripper, generated by visual sensors, simulations, or generative neural networks.



Examples of robotic gripping with three multifingered grippers: Shadow (five fingers), Allegro (four fingers), and Barrett (three fingers).



JULIEN MÉRAND
PhD student



BORIS MEDEN
Research Engineer



MATHIEU GROSSARD
Research Director and Senior Expert

Conventional inverse kinematics (IK) techniques can provide mathematically exact solutions (if any exist) for determining joint configuration from effector (fingertip) position alone. However, because the positions of gripper fingers' intermediate phalanges must also be considered, these techniques often require a posteriori decision-making. For more complex kinematics, numerical approximation algorithms, which may not perform as well in dynamic environments, are used. We developed an innovative machine learning method that leverages a conditional variational autoencoder (CVAE) to reconstruct joint configurations from the robotic gripper's point cloud.

Tests on the MultiDex dataset produced an average joint error of less than 4% with ultra-rapid inference (< 0.05 ms). The algorithm, whose ultra-rapid inference eliminates the need for computationally-costly numerical optimization steps, responds to the demands of real-time environments. Because the training data is generated solely from the URDF model, the method is easy to adapt to any type of gripper.

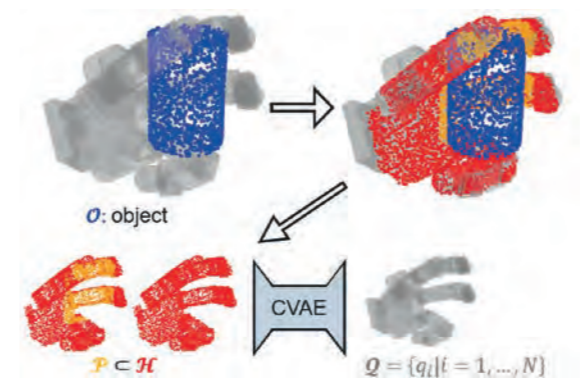
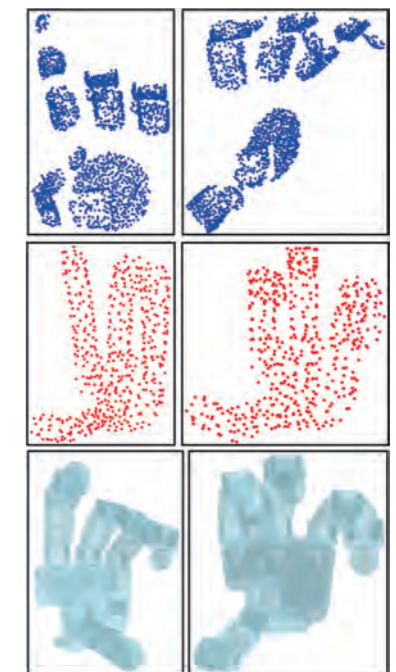


Figure 1. Using this method, we were able to find the joint configuration from a complete or partial 3D point cloud of a multifingered robotic gripper.

4%

4% average joint error with inference time of 0.05 ms (faster than the state of the art) with the MultiDex grip dataset and the Allegro gripper.

Figure 2. Data used: Complete or partial point clouds.



• • • **Flagship publication**

«Leveraging CVAE for Joint Configuration Estimation of Multifingered Grippers from Point Cloud Data», J. Mérand, B. Meden and M. Grossard, 2025 IEEE 21st International Conference on Automation Science and Engineering (CASE), Los Angeles, CA, USA, 2025, pp. 895-900, <https://doi.org/10.48550/arXiv.2511.17276>

Digital instrumentation and the factory of the future

Robotics and interfaces

A cognitive programming interface

Programming robots is a time-consuming endeavor—one that can be an obstacle to their deployment in industrial settings, especially for smaller industrial companies that perform constantly-changing tasks. Robots can be programmed at task level. However, this approach doesn't provide adequate contextualization and requires specialized knowledge. We developed an ontology-based cognitive interface to enable context-sensitive programming that requires no robotics expertise.



With our Cognitive Programming Interface (CPI), non-expert users can program robots by creating a series of sequences consistent with the current state of the scene. The CPI is based on three developments:

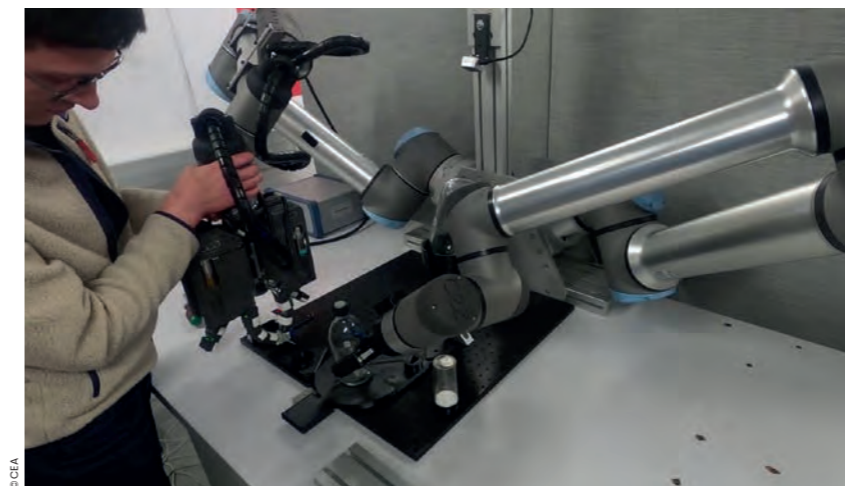
- 1. An ontology** describing a scene, the objects in the scene (such as tables, bottles, or manipulators, which are called actors), their interfaces with the other actors (bottom, gripping area, threading on a canister, etc.), and a set of known skills (gripping, placing, screwing, etc.).
- 2. A context interpreter** updates the state of the world in real time, using changes in states (closing or opening manipulators, etc.) to determine possible interactions between actors based on the interfaces either hidden or left available by each actor.
- 3. A graphical interface** showing the scene in 3D and the skills that can be attained at each stage.

In the CPI's cyclical workflow, the scene is initialized by the world model, which then uses a semantic reasoning engine to calculate the possible skills or abilities. This set of possibilities is sent to the graphical interface and then, according to the actions chosen by the user, the model updates itself. Some steps, such as when a particular relationship or action needs to be specified, may require human intervention.

In the ontological model, actors are described by a geometric primitive, their properties, and, particularly, their interfaces, whose relative positions in the object's frame of reference is known. These interfaces express interaction capabilities (grippable, insertable, placeable, etc.), enabling reasoning not based on objects, but rather on their most likely interactions depending on the context. Relationships between ontology instances are represented as <subject, predicate, object> triplets (e.g. <robot, grasps, object>).

Interaction capabilities are defined by a set of parameters, preconditions, and their effects, which allow relationships to be added, removed, or updated in the ontology. The CPI also allows hybrid queries, which, in the presence of ambiguous or missing information, can call upon either the knowledge base or the user.

A demonstration-based interaction recognition module can automatically deduce the sequence of skills or abilities performed during a kinesthetic demonstration performed manually by the operator. In this module, it is assumed that each skill begins or ends with a change in gripper state. Two main functions—TriggerGrasp and TriggerRelease—interpret movements, contacts, and relationships between objects to update the semantic state of the world. An algorithm then identifies the type of



Training the robotic system on a gesture for a medical use case: The operator manually performs the sequence of operations to be automated (gripping the vial with the gripper on the right and grasping the needle to be inserted into the vial cap with the Abilis dexterous gripper on the left).

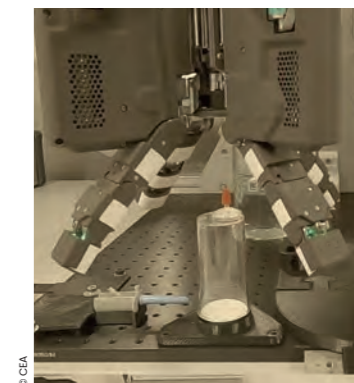


Figure 2: Gripping action

skill using the modified predicates and, finally, deduces the exact parameters by comparing the observed changes with the expected effects of the possible skills.

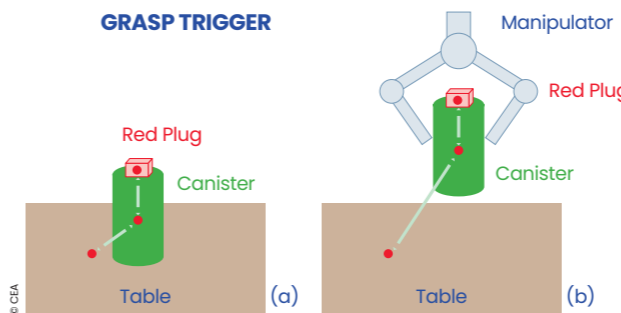


Figure 1. The state of the system before and after input

Flagship publication
 «Cognitive Programming Interface : from Task Level Programming to Coherent Task Level Programming»
 Raphaël Gerin , Julie Dumora , Olivier David , Baptiste Gradoussoff , 2024 IEEE 20th International Conference on Automation Science and Engineering (CASE), Bari, Italy, pp. 1345-1352. <https://doi.org/10.1109/CASE59546.2024.10711577>

EU Tracebot project
 This project addresses the traceability and automation of industrial sterility control processes in the medical and pharmaceutical industries. Two containers placed in racks were handled by two UR10e robots. In the experiment, skills were automatically recognized from a kinesthetic demonstration. Four actions were performed by the operator. First, the two containers were picked up and placed in the rack. Each time the state of the scene changed, the system correctly identified each type of skill and its associated parameters (object, robot, location), comparing the semantic changes observed with those expected. The complete sequence was faithfully reconstructed and displayed via the interface, demonstrating the robustness of the approach. Based on the kinesthetic demonstration, the action can then be repeated automatically by the robot.

Digital instrumentation and the factory of the future

Robotics and interfaces

Active lubrication, a kinesthetic and vibrotactile haptic brake

Active lubrication is an interesting phenomenon. If you take two parts held firmly together by frictional forces, the bond can be loosened by subjecting one of the parts to a high-frequency vibration. Even better, the residual friction force depends directly on the vibration speed, and, due to the system dynamics, the actuator can be switched rapidly. This enables innovative brakes that respond particularly well to the needs of haptic interfaces.

When it comes to the large-scale adoption of haptic devices and, especially, wearables like haptic gloves for virtual reality systems, selecting the right actuator is a major technological challenge. There is currently a lot of research attempting to find an alternative to the traditional bulky electric motors. Recent research has focused on pneumatic, electrostatic, and magnetorheological devices. One thing these devices all have in common is that they are passive actuators (i.e. brakes), with high energy density that enables powerful, yet compact systems. However, they can only work against the user's movement. Nevertheless, in practice, this capability does address a large proportion of useful interactions. CEA-List has been researching active lubrication for a decade, designing a complete integrated actuator that will open the door to a new kind of brake suitable for a wide range of applications.

The principle is simple (Figure 1): A moving part, typically a glass plate, is sandwiched between two piezoelectric actuators. When the actuators are excited to their resonant frequency by an electrical signal, the part is released. The excitation signal's amplitude can then be modulated—very rapidly, at up to 200 Hz or more—to

control the force felt by the user moving the part. In this way, the actuator can render an effort from the kinesthetic to the tactile domain. It is also possible to decouple the two and independently control the force and texture felt.

The design and study of this brake was the subject of Marion Pontreau's PhD research as part of the ANR Wavy project. As a result of this research, friction force was reduced by more than 90%. This impressive improvement demonstrates the usefulness of this approach for haptic

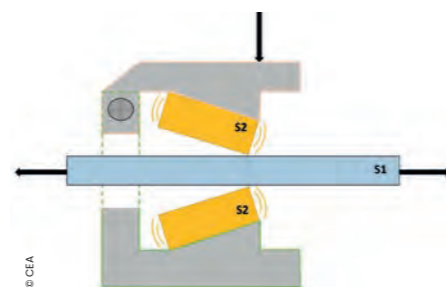


Figure 1. Brake principle: In blue, the part handled by the user. Yellow: Piezoelectric actuators controlling displacement resistance. In gray, the elements that compress the piezoelectric actuators onto the part.



SYLVAIN BOUCHIGNY, Research Engineer and Expert



MARION PONTRÉAU, PhD student



SABRINA PANÉELS, Research Engineer and Expert

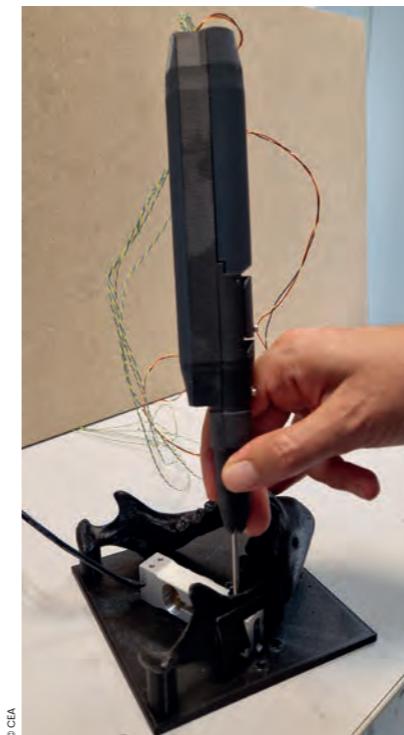
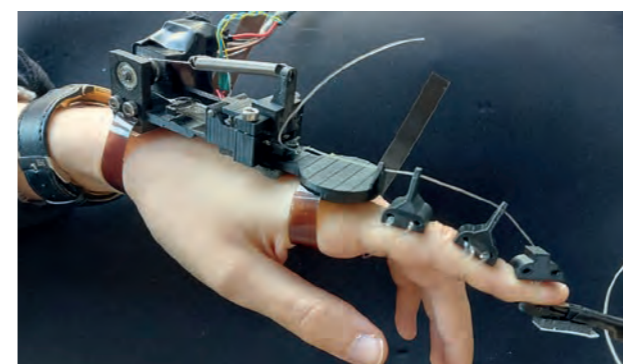


Figure 2. Bone cutter prototype for osteotomy simulation

systems. The actuator developed was tested on various interfaces, including the Wavy haptic glove (see illustration), where it was used to render clamping forces and textures in a demonstration that won the Best Demo Award at IEEE VR 2025. The actuator was also used to design a surgical bone cutter for simulation-based osteotomy training (Figure 2). This research won the Best Student Paper award at World Haptics 2025.



The Wavy haptic glove.

Application
Today, this technology can be applied to several types of interfaces, such as rotary knobs, push-buttons, sliders, and hand-held tools and controllers. It is one of the most affordable technologies on the market.

Major projects
The research was conducted as part of the ANR Wavy and Poc In Labs SATT Paris-Saclay OTACT projects.

Patents
Two patents have been filed, and a third is in the pipeline.
• FR3144344B1 S. Bouchigny, P.H Orefice, M. Pontreau, S. Panéels, "Device for controlling the movement of a part".
• FR3054073A1 E. Leroy, S. Bouchigny, M. Hafez, "Tangible object with improved haptic return and simplified structure and a system comprising at least one such tangible object".

Flagship publications
«WAL: Wearable based on Active Lubrication for Virtual Reality», M. Pontreau, S. Bouchigny, S. Panéels and S. Haliyo, *in Eurohaptics 2024, LNCS, vol. 14769, pp 291-304* https://dx.doi.org/10.1007/978-3-031-70061-3_24
«Design and evaluation of an active lubrication brake for a surgical drilling simulator», M. Pontreau, S. Panéels, S. Bouchigny and S. Haliyo, *in World Haptic Conference, Suwon, South Korea, Jul. 2025 Best Student Paper Award* <https://doi.org/10.1109/WHC64065.2025.11123264>

90% 90%: the typical reduction in friction forces with active lubrication.

Digital instrumentation and the factory of the future

Sensors

Fiber Bragg gratings for ultrasonic measurement at high temperatures

High temperatures are one of the major technological barriers to the deployment of monitoring systems in severe environments. In this research, we demonstrated the first-ever use of fiber Bragg gratings to measure ultrasonic waves at temperatures above 1,000 °C.

Structural health monitoring (SHM) entails acquiring data from the structure being monitored during use. The methods used in SHM have the capacity to provide precise information on the condition of a structure, making maintenance more efficient. The instruments used to collect data must be able to do so without disrupting operations. They must also be able to withstand the rigors of the structure's operating environment. To address these challenges, CEA-List is developing guided-ultrasonic-wave-based systems. In these systems, waves highly sensitive to disturbances in the structure being monitored are propagated over long distances. Our approach is to use fiber Bragg grating sensors to measure the waves. Due to the small fiber diameter, this solution does not disrupt normal use of the structure. In addition, the sensors can withstand harsh conditions like extreme temperatures, ionizing environments, and strong electromagnetic fields.

Fiber Bragg grating sensors also offer the advantage of being very sensitive to temperature changes and to strain. Compared to the standard measurements taken with this type of sensor, the ultrasonic waves used in our approach are characterized by small, high-frequency deformations, which require a specific interrogation technique called edge filtering. We developed a sensor interrogator capable of measuring ultrasonic waves despite variations—wider amplitude and lower frequency—due to changes in environmental and operating conditions. The system was used for a measurement campaign at the Andra Underground Research Laboratory in Bure, France.

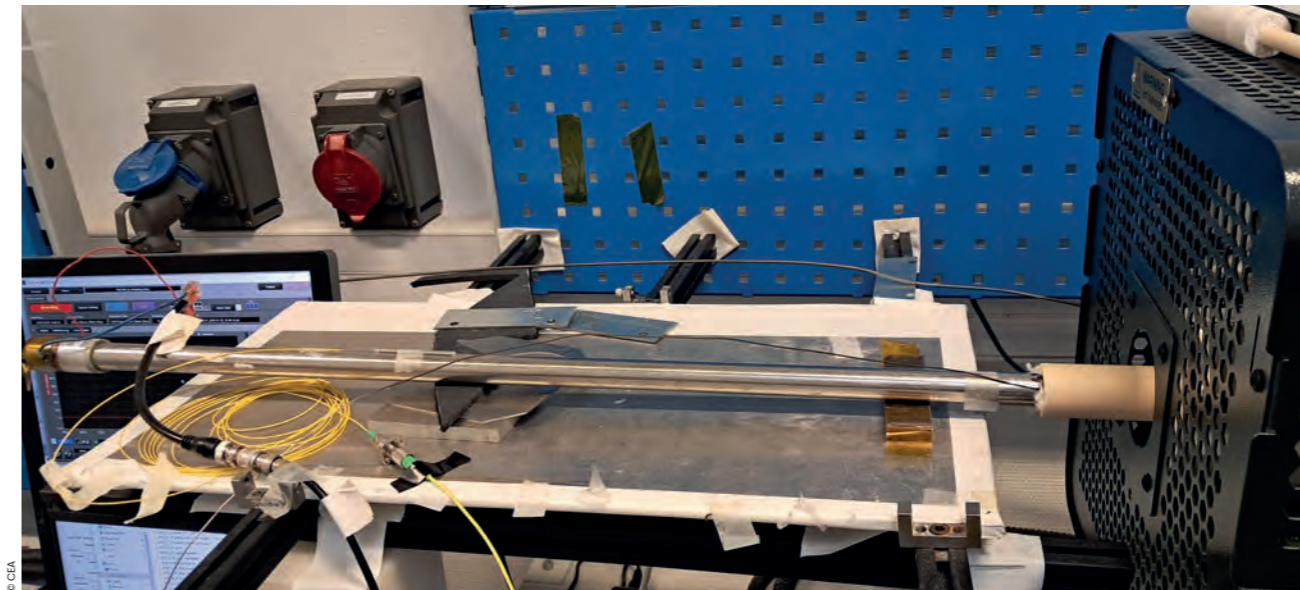
In one experiment, a sensor was mounted on a metal rod inside a furnace to test the system's measurement capabilities in severe environments. The rod was heated to over 1,000 °C. An ultrasonic source, incapable of withstanding such high temperatures, was attached to the end of the rod extending outside the furnace. Ultrasonic signals were effectively measured using the sensors, a first at such high temperatures. This experiment demonstrated the feasibility of using guided-ultrasonic-wave-based monitoring systems in severe environments.



ANTOINE GALLET
Research Engineer



ARNAUD RECOQUILLAY
Research Engineer and Expert



High-temperature Bragg grating measurement experiment. The metal rod and optical fiber are inserted into the furnace on the right, and a piezoelectric transducer, on the left, emits ultrasound into the rod.

The research, which is ongoing, is now focusing on improving the measurement and, specifically, the system's ability to interrogate a larger number of sensors (currently limited to seven) on the same fiber simultaneously.



Figure 1: Optogéro measurement system deployed at Bure



Figure 2: Optical fiber for integration into a system

Guided-ultrasonic-wave measurement at temperatures above 1,000 °C

1 000 °C

« Fiber optic sensors enable non-intrusive monitoring in severe environments. »

•• ARNAUD RECOQUILLAY

•• Application
High-temperature monitoring (furnaces, aeronautics, nuclear, etc.)

•• Project
Euratom FIND

•• Flagship publications
«Ultrasonic Guided Waves Measurements using Bragg Gratings in Optical Fibers under Varying Environmental Conditions». A. RECOQUILLAY, N. ROUSSEL, L. MAURIN, et al. In: 50th Annual Review of Progress in Quantitative Non Destructive Evaluation. American Society of Mechanical Engineers, 2023. p. V001T09A009. <https://doi.org/10.1115/QNDE2023-118544>



«Ultrasonic waves detection above 1000 degree Celsius with Fiber Bragg Grating sensors». A. GALLET, A. RECOQUILLAY, N. ROUSSEL, et al. In: 2025 IEEE Sensors Applications Symposium (SAS). IEEE, 2025. p. 1-5. <https://doi.org/10.1109/SAS65169.2025.11105178>

Digital instrumentation and the factory of the future

Sensors

Quantum sensors lift the veil on neutrinos

The BeEST project consortium has been using ultra-sensitive quantum sensors to study virtually undetectable particles called neutrinos. Their efforts resulted in the first-ever measurement of a neutrino's wave packet upon its creation. The lithium-7 recoil energy caused by the decay of beryllium-7 was analyzed, revealing that the neutrino extends physically over a much wider area than the nucleus—challenging our previous understanding of weak nuclear decays.

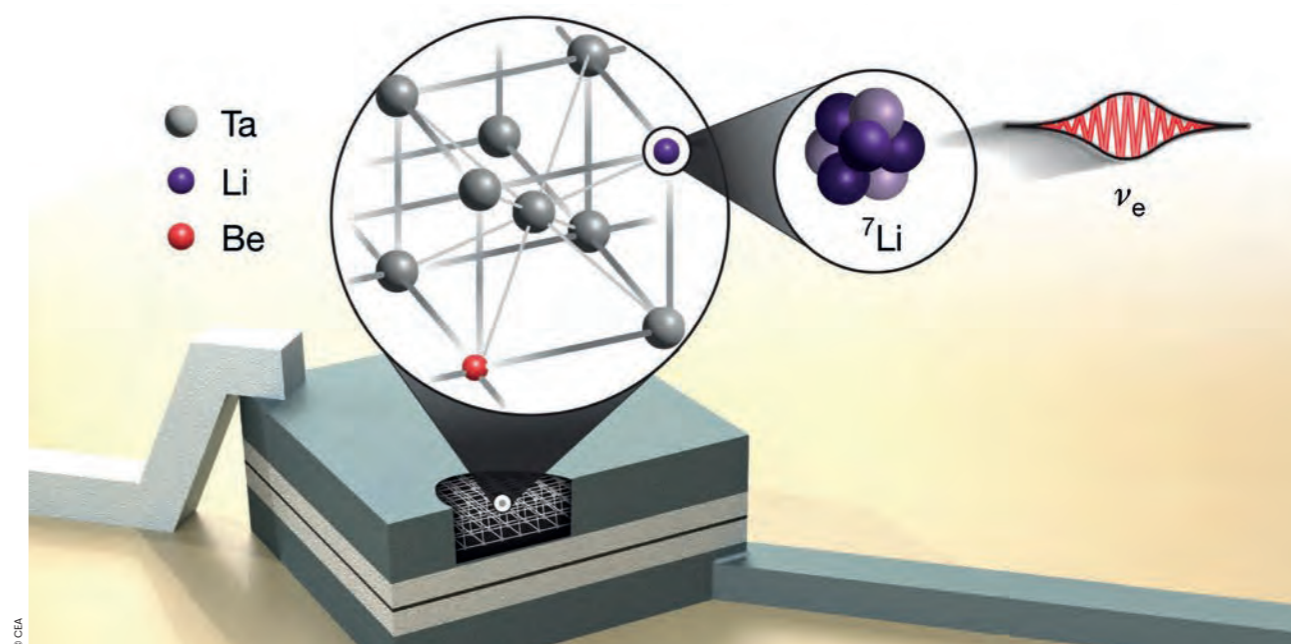
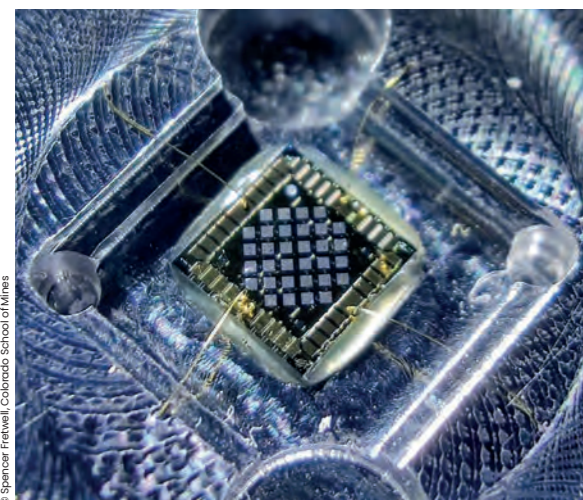


Figure taken from Nature 638, 640–644 (2025)



Set of 32 quantum sensors integrated on a chip and irradiated to implant beryllium-7. Credit: Spencer Fretwell, Colorado School of Mines

The existence of neutrinos was first hypothesized by Wolfgang Pauli in 1930 in a desperate attempt to save the fundamental principle of the conservation of energy in beta decay, in which it appeared that energy was lost. It turned out that neutrinos do exist, but their interaction with matter is extremely limited. Studying neutrinos requires very large detection systems like the IceCube Neutrino Observatory in Antarctica, with its cubic kilometer of ice. Today, still, neutrinos' properties remain mysterious. While the Standard Model of particle physics predicts neutrino masses to be zero, oscillation experiments have shown that they do indeed have mass. Three types of neutrinos are currently known, and the existence of a fourth, the sterile neutrino, coupled only to gravity, is still merely hypothetical.

The BeEST consortium, led by Kyle G. Leach, was created in 2020 to study neutrinos. LHNH has been involved in the project since its inception, contributing to an innovative "tabletop" experiment to measure the properties of neutrinos indirectly using extremely small sensors (about the diameter of a human hair). The radioactive beryllium-7 nucleus, which decays by capturing an atomic electron, is implanted in a superconducting tunnel junction—a quantum sensor with exceptionally high energy resolution. The ultra-precise measurement of the nucleus' recoil energy—a spectrum of just a hundred electronvolts—provides information about the neutrino emitted, as the two particles are quantum-entangled.

Neutrinos, like any quantum particle, cannot be specifically located. However, there is a probability that one will be present in a certain area of space. Until now, this probability was assumed to correspond to the size of the neutrino's nucleus. The BeEST experiment yielded a very unexpected observation: As soon as the neutrino is created, it extends over an area thousands of times larger than the nucleus. This finding upends everything we thought we knew about radioactive decay.

The method in the experiment is so precise that it can also be used to observe the effect of the chemical environment on the nuclear process. One of the BeEST consortium's first findings shows that, between measurement in a tantalum detector and a theoretical prediction for a nucleus in a vacuum, the probability of beryllium-7 electron capture can vary by a factor of two. The BetaShape code, developed by LHNH over the past fifteen years and now an international reference for the evaluation of beta decay data, was used for the theoretical prediction. The effect of the chemical environment is currently the main scientific bottleneck limiting the accuracy of the physical information that can be extracted from the BeEST experiment. Additional—and complex—theoretical research to model this effect as accurately as possible is in progress with Paul-Antoine Hervieux, Professor at Strasbourg University.

Project

The quantum sensors' exceptional sensitivity will open up new possibilities for the study of the chemical environment's effect on nuclear processes. A project funded through an ERC Synergy Grant currently in progress is focusing on developing a multi-scale model that preserves quantum hybridization, from the material (10–3 m) to the nucleus (10–15 m). This model would remove the Standard Model physics component from the beryllium-7 spectrum. While neutrinos cannot be considered dark matter according to the Standard Model, this advance could enable the exploration of a new physics—that of massive neutrinos—potentially putting the particles back in the running as candidates for dark matter. It could also lead to a new approach to predicting the best materials to use to influence the radioactive decay process in a controlled way.

Flagship publication



This breakthrough was published in the journal *Nature*:
«Direct experimental constraints on the spatial extension of a neutrino wavepacket», J. Smolksy et al.,
Nature 638, 640–644 (2025),
<https://www.nature.com/articles/s41586-024-08479-6>

“ The BeEST experiment provided a new perspective that has broadened our understanding of radioactive decay. ”

••• XAVIER MOUGEOT

Digital instrumentation and the factory of the future

Sensors

Electrochemical measurement techniques for nitrite detection

CEA-List is innovating with diamond—an exceptional material that presents remarkable electrochemical properties—to detect nitrites in urine samples. This advance could aid in the early detection of urinary-tract infections.

Biological testing laboratories and hospitals commonly test for nitrites to help diagnose urinary-tract infections (UTIs). When a sample tests positive for nitrites, it could mean that a Gram-negative bacterium—most often *Escherichia coli*, responsible for 75% to 95% of UTIs—is the culprit. Nitrites show up in urine when endogenous nitrates are converted into nitrites, indirectly indicating an infection caused by these types of bacteria. The most used rapid-nitrite-testing method involves Greiss-reaction-based test strips that change color due to the interaction of nitrites with Greiss reagent. This testing method is fast, inexpensive, and doesn't require any specialized knowledge. However, the interpretation of the results can be subjective, and the test's sensitivity is relatively limited (to around 10 μM in practice). Plus, it doesn't provide an indication of the actual concentration of nitrites present in the sample.

CEA-List has been working to develop an alternative nitrite-testing method with Usense, a France-based startup developing a connected wearable medical device that accurately and instantaneously measures several biomarkers in urine.



EMMANUEL SCORSONE
Research Engineer,
Senior Expert



GUILLAUME LEMETAIS
CEO, Usense



Figure 1: Boron-doped diamond electrodes produced at CEA-List using microwave plasma-assisted chemical vapor deposition.

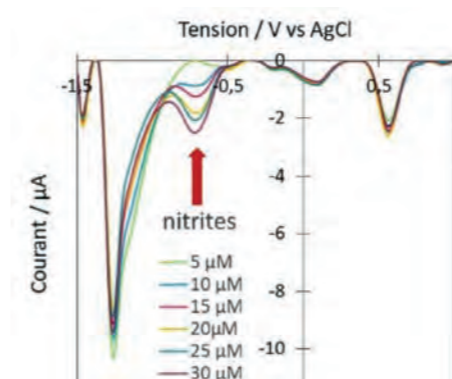


Figure 2: A voltammogram, recorded during the process development phase, showing a boron-doped diamond electrode's response to a labeled urine sample containing nitrites.

The partnership resulted in an analytical method using boron-doped diamond electrodes to electrochemically detect nitrites in urine. The remarkable properties of the boron-doped diamond electrodes fabricated at CEA-List are at the center of this new method. Diamond offers a wide potential window in aqueous media, low double-layer capacitive current, and a self-cleaning electrode surface that leverages an electrochemical process patented by CEA-List several years ago. This self-cleaning capability means that tests can be repeated many times—tens of thousands according to our estimates—without having to replace the electrodes. The underlying principle is the electrochemical reduction of nitrosamines, which are formed in urine as a result of the spontaneous reaction of nitrites with tryptophan from the natural breakdown of proteins in the body.

The method was validated on hundreds of urine samples in biological testing laboratories and in hospitals in partnership with Usense. The detection of nitrites in urine is very selective, with a threshold of around 0.5 micromolar.

«Diamond, with its unique physico-chemical properties, hasn't lost its sparkle, bringing new capabilities to a medical use case with strong potential to impact society.»

••• GUILLAUME LEMETAIS



••• Use cases, applications, technology transfer

CEA-List and Usense are currently in talks to transfer the diamond-electrode-based nitrite testing technology. In the short term, Usense would like to integrate the technology into its Jimini urinalysis device, which is designed to replace strip tests and provide a more reliable, sensitive, and digitizable alternative to conventional testing. Since 2023, pharmacists in France have been allowed to perform urine tests eligible for reimbursement from the national healthcare system. These in-pharmacy tests, intended to diagnose symptoms associated with acute uncomplicated cystitis in female patients, must include nitrites and leukocytes.

••• Patent

The process for detecting nitrites in urine using diamond electrodes was patented in 2024.

••• Major partnerships

Usense



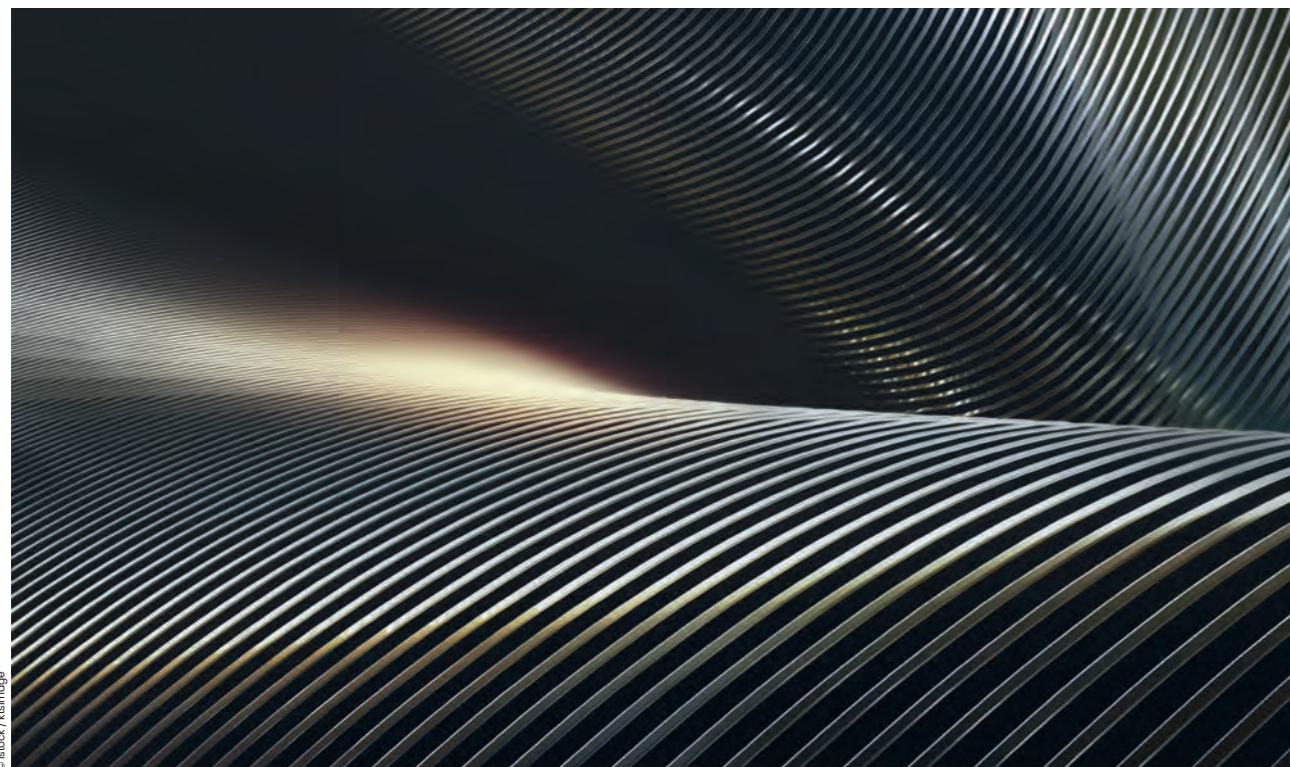
Digital instrumentation and the factory of the future

Modeling and characterization

Continuous eddy-current monitoring of fatigue degradation in an unnotched steel sample



ANASTASSIOS SKARLATOS
Research Engineer and Senior Expert



Metal fatigue, which can result in the gradual degradation of a structure's mechanical properties, is a critical phenomenon in a wide range of industrial settings. To ensure the safety and durability of metal structures, the ability to monitor degradation is vital. Electromagnetic monitoring methods, with their sensitivity to microstructural changes, are particularly well-suited to this task.

We studied a C22E/C22R-grade structural steel sample with a minimum cross-section of 7.6 mm and a gauge of 12 mm. We applied a cyclical load of 230 MPa at a frequency of up to 5 Hz until failure. To monitor variations in magnetic permeability, eddy-current measurements were taken at different excitation frequencies (50 kHz, 100 kHz, 250 kHz, and 500 kHz).

A semi-analytical model was used to solve the direct problem: the calculation of impedance as a function of magnetic permeability. The experimental results were compared to simulations to validate the model. The material's relative permeability, μ , was estimated in the range 150-180, with permeability decreasing as a function of time over the course of the experiment.

The results showed a clear difference between the low-frequency (below 3,000 fatigue cycles) and high-frequency portions of the measurements. This difference can be attributed to different microstructural mechanisms occurring at different times during the experiment. The high-frequency portion of the measurements revealed almost linear behavior, confirmed by an analysis of the actual voltage values.

We also took magnetic Barkhausen noise (MBN) measurements to round out the eddy-current measurements. Thermographic data was used for correlation. We also factored in temperature variations to correct for any drift caused by the sample heating up.

According to our results, eddy-current measurements are suitable for monitoring fatigue-induced degradation in metal. Variations in magnetic permeability can be correlated with microstructural changes, enabling the non-destructive monitoring of the material's condition.

Now that eddy-current measurements have been demonstrated on the monitoring of fatigue-induced degradation, the door is open to new industrial solutions for real-time monitoring to help improve the safety and durability of metal structures.

Major flagship

«Continuous eddy-current monitoring of fatigue degradation in an unnotched steel specimen». Skarlatos A, Boller C, Lyamkin V, et al. *International Journal of Applied Electromagnetics and Mechanics*. 2025; <https://doi.org/10.1177/13835416251345615>

Use cases, applications, technology transfer

The continuous monitoring of metal structures for fatigue-induced degradation is crucial to ensuring safety and durability. Eddy-current measurements used in conjunction with semi-analytical models offer an effective method for assessing microstructural changes and monitoring the condition of the material. These results confirm that our approach can be used for real-time monitoring, creating new possibilities for industrial solutions.

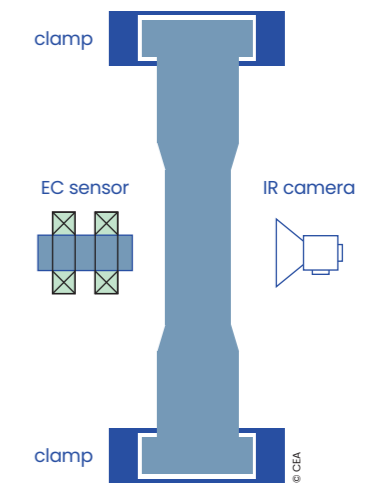


Figure 1: Sensor assembly.

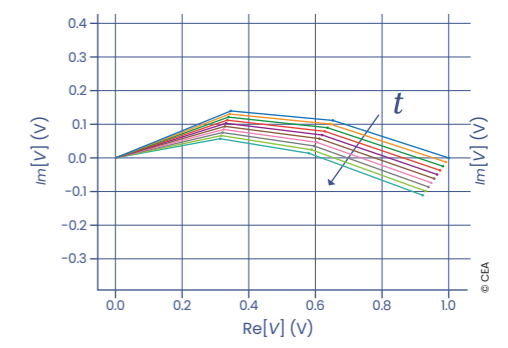


Figure 2: Time evolution of signatures in the complex plane.

Digital instrumentation and the factory of the future

Modeling and characterization

AI algorithms for gamma-ray spectrometry used in field measurement

In gamma spectrometry, the demand for automated algorithms for rapid decision-making in low-statistics in situ measurements is high. However, speed is not the only consideration. Algorithms must also be compatible with detectors with low energy resolution, and false alerts must be managed robustly, especially if the spectrum is distorted.

CEA-Leti combined machine learning and statistical techniques in SEMSUN, a hybrid spectral demixing algorithm. An interpolation autoencoder (IAE) was used to learn spectral distortions induced by Compton attenuation and scattering phenomena in the source's environment. The autoencoder, a neural network, can extract the characteristics of a spectrum in a compressed representation (encoder) and can be used as a generative model to reconstruct the spectra from the learned characteristics (decoder). The spectral unmixing technique is based on a mixing model of spectral signatures specific to each radionuclide present. SEMSUN uses the IAE model to impose a constraint on the deformation of these spectral signatures in solving an inverse problem. This hybrid approach optimizes radionuclide identification and counting by explicitly integrating the physical characteristics of a measurement, unlike a state-of-the-art approach that relies on end-to-end learning.

The IAE model training was based on Geant4 code simulations of radiation-matter interactions in a geometric configuration in which a point source is placed at the center of a sphere (steel, lead), with a 3"×3" NaI(Tl) detector (Figure 1). Spectrum reconstruction results show that the IAE model effectively captures spectral deformation, allowing us to establish the proof of concept^[1].

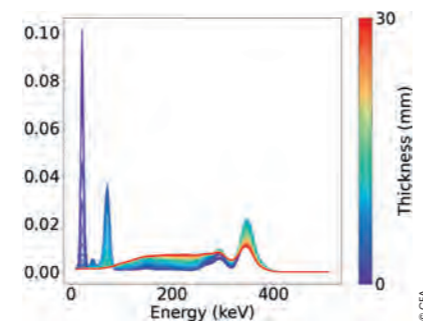


Figure 1: Spectral deformation of a Ba-133 source placed at the center of a steel sphere, as a function of sphere thickness (Geant4 simulations).



CHRISTOPHE BOBIN
Research Engineer
and Senior Expert

To automatically identify radionuclides, SEMSUN was combined with a model selection technique based on a likelihood ratio test, resulting in a new algorithm, MoSeVa^[2]. A sparsity criterion was applied to the various spectral signatures grouped together in a library to determine the presence of a radionuclide. Tests were carried out on different natural background radiation mixtures (up to four radionuclides in a library of twelve spectral signatures). Different counts in the spectra were also studied up to a minimum count of 2,500 events (Figure 2).

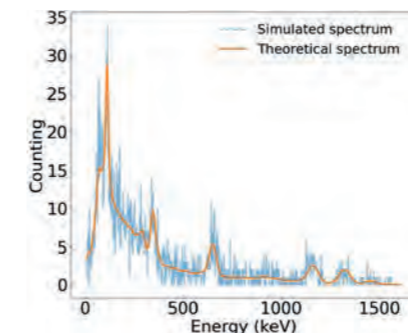


Figure 2: Natural background spectrum of Co-57, Co-60, Ba-133 and Cs-137 simulated for a total of 2,500 events.

Hybrid spectral unmixing was compared to a state-of-the-art end-to-end learning method based on convolutional neural networks used as classifiers or for regression applications. The results confirm that hybrid spectral demixing performs better on low-statistics measurements in terms of identification and robustness, with the expected false alert rate. Hybrid spectral demixing also enables more accurate counting beyond low statistics^[3]. The GammaBench benchmark (including spectra, evaluation metrics, etc.) can be used to compare the performance of new developments with hybrid spectral demixing.

**×10
MORE
ACCURATE**

Improvement using hybrid spectral demixing on quantization compared to end-to-end methods.

“Hybrid spectral demixing offers a solution to the problem of low-statistics in situ measurements sensitive to spectral variability.”

• • • CHRISTOPHE BOBIN

• • • Use cases, applications, technology transfer
In situ environmental measurements following nuclear or radiological accidents involving the release of radionuclides, detection of illicit trafficking of radioactive materials, radiological characterization during nuclear facility remediation and decommissioning, etc.

• • • Flagship publications



[1] «A hybrid Machine Learning unmixing method for automatic analysis of gamma-spectra with spectral variability». Phan D.T., et al., *Nuclear Inst. and Methods in Physics Research A* 1060 (2024) 169028.
<https://doi.org/10.23919/EUSIPCO63174.2024.10715033>



[2] «Automatic identification and quantification of γ -emitting radionuclides with spectral variability using a hybrid Machine Learning unmixing method». Phan D.T., et al., *Radiation Physics and Chemistry* 232 (2025) 112654.
<https://doi.org/10.1016/j.radphyschem.2025.112654>



[3] «Comparative study of machine learning and statistical methods for automatic identification and quantification in γ -ray spectrometry». Phan D.T., et al., *Nuclear Inst. and Methods in Physics Research A* 1083 (2026) 171088.
<https://doi.org/10.48550/arXiv.2508.08306>

Digital instrumentation and the factory of the future

Modeling and characterization

Photoneutrons and photofission, scientific weapons against trafficking

The ability to effectively inspect maritime containers is crucial to European security. CEA-List drew on more than three decades of experience with active non-destructive nuclear measurement to develop two groundbreaking methods to improve the detection of nuclear material and of drugs.

CEA-List researchers are currently developing two major innovations in active non-destructive nuclear measurement methods to fight the trafficking of hazardous materials and illicit substances. The first combines a linear electron accelerator (linac), photoneutron spectrometry, and artificial intelligence to identify light elements characteristic of explosives and narcotics. The second, developed for the EU MULTISCAN 3D project with Munich's LMU^[1] -CALA^[2], is laying the foundations for a breakthrough in nuclear material detection technology.

The method developed to detect illicit substances is based on a linac that generates a braking radiation capable of inducing photonuclear reactions. Liquid scintillators are used to detect the neutrons emitted, whose spectra contain the signatures of the elements (carbon, nitrogen, oxygen) present. An intense burst of photons allows the neutrons of interest to be identified. The spectra, despite their rich structures, are often distorted by the geometry of objects inspected and the detector's response, a challenge that is effectively addressed by automated analysis. The DeepNSI deep learning model, based on convolutional neural networks trained by Monte Carlo simulation, is able to identify elements even at very low concentrations (less than 4% for nitrogen). Other algorithms, like Nonnegative Elastic Net, can also be used to estimate the relative concentrations of the different elements present. The resulting chemical identification is advanced, robust, and fast enough to meet the requirements of freight inspection.

CEA-List also reported a world-first experimental demonstration of photofission on impoverished uranium using a totally new photon source, marking a major advance in the development of tools to fight the illegal trafficking of nuclear material. The photon source developed is based on Inverse Compton Scattering (ICS): A femtosecond laser beam accelerates high-energy electrons in a plasma, producing photons of sufficient energy to initiate photofission. This is the first time this type of photon source has been used in place of a



VALENTIN BLIDEANU
Research Engineer
and Research Director



ADRIEN SARI
Research Engineer,
Senior Expert

traditional linac. The laser's quasi-mono energy, integrability, and scalability clear a major new path toward the future deployment of active nuclear measurement to detect actinides in shipping containers.

While these two advances focus on the detection of different targets (drugs and nuclear material), both are a testament to the emergence of disruptive, more sensitive, and smarter technologies capable of making Europe's ports much more secure.

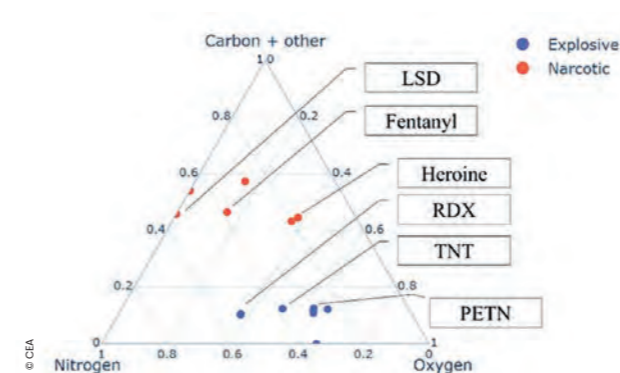


Figure 1: Ternary graphs for several molecules of interest as a function of their carbon, oxygen, and nitrogen ratios based on the photoneutrons in the total spectrum.

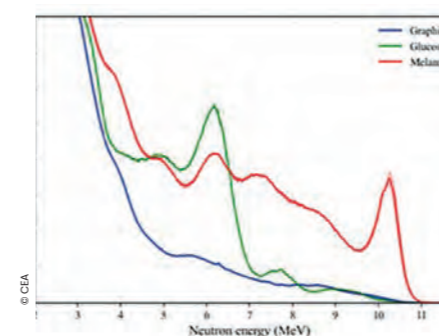


Figure 2: Photoneutron spectra measured for three materials: graphite (carbon-rich), glucose (oxygen-rich), and melamine (nitrogen-rich), enabling the specific signatures of these elements to be identified through the structures present.

“ AI analyzes distorted photoneutron spectra and extracts the elementary signatures that cannot be isolated by spectrometry alone”

• VALENTIN BLIDEANU

“ Our advance using a laser-based ICS source marks a major step toward the development of advanced technologies for the detection of nuclear material.”

• ADRIEN SARI

• Use cases, applications, technology transfer
DeepNSI marks an advance toward the identification of light elements in complex photoneutron spectra for the detection of illicit materials, radiation protection, and decommissioning. CEA-List is currently exploring opportunities to transfer the photoneutron spectrometry technology to an industrial partner. The new laser-plasma source will lead to research on new photofission regimes.

• Patents
• IPA-SN (Dispositif et procédé de détection d'une substance particulière dans un objet par interrogation photonique active): French patent no. **FR2112182**
• DeepNSI (Méthodes d'analyse avancées des spectres photoneutroniques): French patent no. **FR2210022**

• Major project and/or partnership
EU MULTISCAN 3D project

• Flagship publication
“DeepNSI: Element identification in experimental photoneutron spectra for illicit material detection”, C. Besnard-Vauterin, V. Bliideanu, B. Rapp, *Applied Radiation and Isotopes* 225 (2025) 112014; article on the laser-plasma source forthcoming. <https://doi.org/10.1016/j.apradiso.2025.112014>

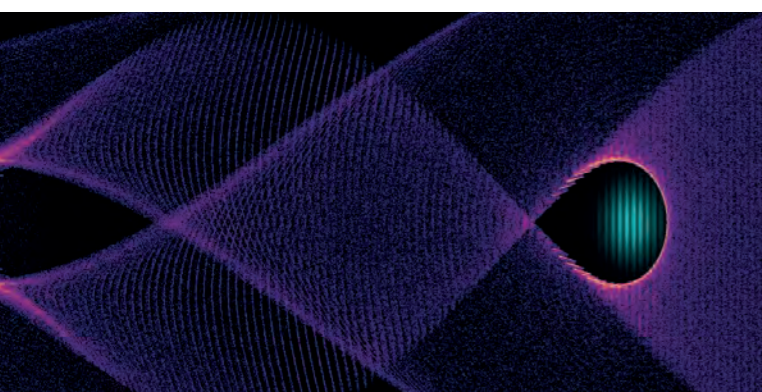


Illustration of the laser-plasma acceleration process.

[1] LMU: Ludwig-Maximilians-Universität München
[2] CALA: Center for Advanced Laser Applications

According to the European Drug Report 2024, nearly 70% of seizures made by EU customs officials take place in ports, mainly in maritime containers.

70%

05

Cybersecurity and trustworthy digital technology

Software is everywhere, and it is becoming increasingly complex. This, and advances in quantum computing, are driving growing demand for new cybersecurity solutions. Building reliable systems capable of detecting and prioritizing vulnerabilities and developing cryptographic countermeasures will require approaches that span theory and operational realities.

Amid the increasing complexity of today's software, guaranteeing trust in an application's ability to withstand attacks presents some major challenges. Formal-methods-based static analysis can mathematically verify the absence of entire classes of vulnerabilities—like buffer overflows—commonly exploited by attackers. In traditional approaches, a full analysis must be completed every time the code is modified—a requirement that is incom-

patible with continuous integration workflows. More recent formal static analysis methods are better aligned with the agile development of critical systems. By reusing previous results, these incremental analysis methods drastically reduce analysis times without sacrificing the necessary mathematical guarantees.

The problem is that the automated detection of vulnerabilities now far outpaces our ability to correct them. The Linux kernel, due to its size, complexity, and the amount of code it contains, is associated with thousands of vulnerabilities in CVE databases each year. Given the sheer number of vulnerabilities, it is impossible to know which ones should be given priority. And Linux is far from an isolated case. The answer lies in vulnerability exploitability assessment, which looks at the likelihood that an attacker can effectively take advantage of a vulnerability. New techniques can be used to analyze the exploitable values of different parameters, enabling the automated classification of truly critical vulnerabilities, validated on real CVEs.

Finally, advances in quantum computing are raising pressing questions about today's encryption methods. The risk of attackers collecting encrypted data today and cracking it tomorrow with quantum computers is very real. Post-quantum cryptography and fully-homomorphic encryption offer future-proof protection, made practical by hardware acceleration, with performance gains of up to 94% to 96%.



86 Vulnerability detection

- 86 Incremental software verification for cybersecurity
- 88 Exploitability by attackers and vulnerability prioritization

90 Security-enabling technologies

- 90 Fast fully-homomorphic encryption
- 92 Hardware accelerator tightly coupled to a RISC-V core for post-quantum cryptography

Cybersecurity and trustworthy digital technology

Vulnerability detection

Incremental software verification for cybersecurity

In cybersecurity, newly-introduced software vulnerabilities must be detected immediately. The generic incremental analysis techniques recently integrated into Frama-C can verify code quickly after local modifications. The early results are very encouraging in terms of analysis times, while reliability remains excellent.

Tools that simply detect vulnerabilities in source code cannot ensure that there are no vulnerabilities in the code at all. This requires a different approach.

Frama-C, CEA-List's formal verification tool, addresses this challenge by using mathematical proof to automatically verify the absence of a broad class of vulnerabilities like buffer overflows, which, due to insufficient precautions during programming, allow the amount of data to exceed the memory's storage capacity.

This kind of verification does come at a cost. However, it is much better to do it early in the development cycle when any vulnerabilities introduced into the code are still easy to correct. And, if developers are aware of the demands of early and automated formal verification, they can adapt their programming style accordingly—an added benefit. Ideally, Frama-C is used as part of a continuous integration process, a development practice in which code is tested and validated with each modification, right from the start of the project. However, for Frama-C to be effective in this context, its analysis time must be similar to the interval between two code modifications.

We introduced incremental analysis techniques into Frama-C to meet this requirement. Instead of redoing a full analysis after each code modification, the previous analysis is reused as much as possible to speed up the next analysis. The major challenge is to formally prove that the reuse of previous analysis results doesn't compromise the correction of any vulnerabilities. In short, the incremental analysis must be as reliable as the full analysis.

We used two techniques to achieve this.

The first technique involves reusing code function summaries, which correspond to mathematical properties deduced from each function, produced by the first analysis. By comparing code versions and exploring their structure, we can determine which summaries can be reused for the new analysis of a function.

The second technique speeds up the search for loop invariants (logical properties of the code). Invariants discovered during a previous analysis are reused as starting points so that invariants can be found more quickly in the new version of the code.



© CEA
VALENTIN PERRELLE
Research Engineer



© CEA
MAMY RAZAFINTSIONINA
PhD student



© CEA
JULIEN SIGNOLES
Research Director



« When used early in software development, Frama-C enables vulnerabilities to be detected while they are still easy to correct. »

••• VALENTIN PERRELLE

Initial experiments on open-source code with a public modification history showed that incremental code analysis with Frama-C reduces analysis time by a factor of 1.5x to 28x, with an average improvement of 7.83x. The experiments also revealed very little variation in the analysis results, with no negative impact on reliability. These techniques are, therefore, compatible with implementation in continuous integration workflows.

••• **Flagship publications**

This research, conducted in partnership with Sorbonne University as part of the SecurEval project, part of the national PEPR Cybersecurity program funded by France 2030, resulted in one publication:



M. Razafintsialonina, D. Bühler, A. Miné, Valentin Perrelle, J. Signoles «**Reusing Caches and Invariants for Efficient and Sound Incremental Static Analysis**» at ECOOP 2025 (rang A)
<https://doi.org/10.4230/DARTS.11.2.15>



and «**Réutilisations de caches et d'invariants pour l'analyse statique incrémentale**» in JFLA 2024
<https://inria.hal.science/hal-04406463>

Figure 1: Reuse of part of the previous analysis to analyze modified code.

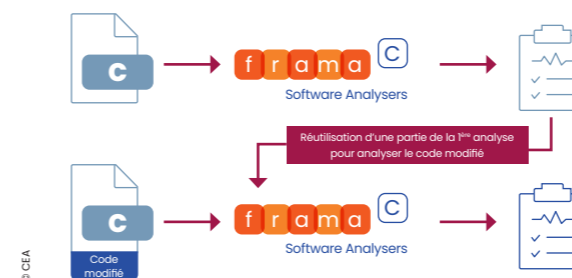
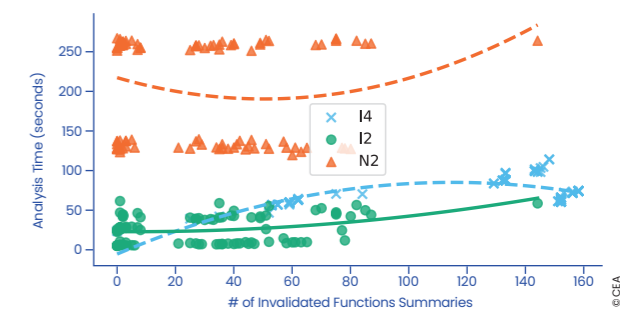


Figure 2: Time saved by incremental analysis (green) compared to full analysis (orange) on one hundred open-source cryptographic codes, with the number of non-reusable function summaries on the x-axis.



Cybersecurity and trustworthy digital technology

Vulnerability detection

Exploitability by attackers and vulnerability prioritization



The automated detection of vulnerabilities and, especially, automated code analysis, has been one of cybersecurity’s greatest successes of the past fifteen years. Fuzzing, symbolic execution, and abstract interpretation are among the software testing techniques that have made the identification of vulnerabilities faster than what we previously could have imagined. The ability to uncover vulnerabilities at unprecedented rates does, however, raise the question of what to do about the very large number of vulnerabilities identified.

Google’s OSS-Fuzz, which is just one of many public fuzzing platforms, contains tens of thousands of bugs—a number most development teams cannot even hope to keep up with. But the sheer number of vulnerabilities is not the only challenge. The reports generated by automated vulnerability detection tools provide little information about the security impacts of the vulnerabilities detected. This makes it practically impossible to prioritize the most dangerous vulnerabilities so that they can be corrected quickly. The need for automated vulnerability and exploitability testing techniques has never been greater.

“ Today’s verification tools identify far more vulnerabilities than what can be corrected. So, how do you know where to begin? By using formal methods, vulnerabilities can be prioritized according to how dangerous they are.”

••• SÉBASTIEN BARDIN

Our research introduces a formal expression of the exploitability of a given vulnerability to tell us how likely it is that an attacker will be able to influence the behavior of a certain kind of vulnerability through program inputs. Intuitively, a vulnerability whose effects on a program can be finely controlled by an attacker is more dangerous than a vulnerability that is less exploitable. Our formal evaluation technique measures the number of unique behaviors that an attacker can control. But knowing the number of behaviors is only part of the solution. You must also be able to determine how dangerous the behaviors are. Our approach is to include all this information to arrive at a priority score—a unique and complex challenge. Formal methods ensure that the results are transparent and explainable. However, they can also be used to explore areas that have only rarely been addressed by researchers until now. And, in the limited research that does exist, the methods used always rely on either an expert’s judgment or on a rough classification that assumes all vulnerabilities in a class are equally dangerous.

Our research was published in the proceedings of Usenix Security 2025, one of the world’s leading scientific conferences on security (A-rank*). It also resulted in a



new plug-in for CEA-List’s BINSEC binary code analysis platform. We tested and validated our approach on verification tasks involving real memory corruption vulnerabilities taken from CVEs (Common Vulnerabilities and Exposures databases) and on a benchmark of vulnerabilities detected by Google’s OSS-Fuzz. Our assessment covered a wide range of programs from widely used libraries like libxml2 and libpng, large-scale complex systems like openssl, and the universal boot-loader u-boot, for example. Our approach produced a relevant classification of the danger of different vulnerabilities, establishing a new state of the art.

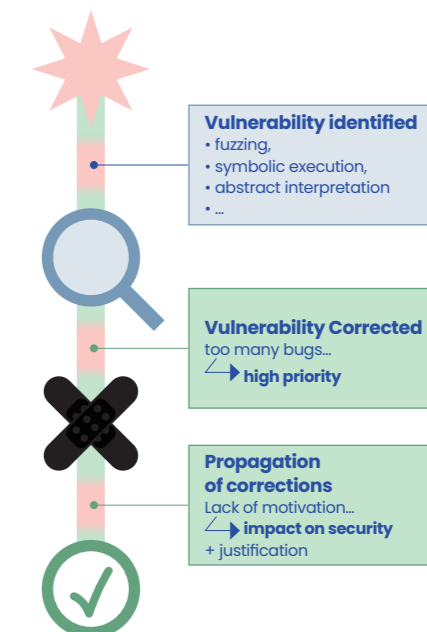


Rise in the number of vulnerabilities detected (and still open) in Linux using fuzzing, a random testing technique.

••• Major project and/or partnership
 PEPR Cybersecurity/REV. Interest from ANSSI and Schneider Electric.

••• Flagship publication
«Attacker Control and Bug Prioritization».
 G. Lacombe et S. Bardin. *USENIX Security 2025*.
Flagship conference in system security (A)*
<https://doi.org/10.48550/arXiv.2501.17740>

••• Determining the degree to which an attacker can effectively control vulnerabilities is only the first step toward automated vulnerability detection. Among the many problems that must still be solved are how to assess other aspects of vulnerability exploitability, how to combine heterogeneous analysis results into a single score and ensure that the score is relevant, and, finally, how to deploy these analyses on a large enough scale with computing speeds fast enough for on-the-fly testing. This research marks a foundational first step. We are continuing to explore the extensions of these concepts and their application to real-world use cases in research for the national PEPR Cybersécurité initiative.



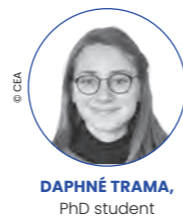
Our method for prioritizing vulnerabilities

Cybersecurity and trustworthy digital technology

Security-enabling technologies

Fast fully-homomorphic encryption

In fields like defense, industry, healthcare, and finance, data privacy is a major challenge. One promising avenue is using fully-homomorphic encryption (FHE) to compute on encrypted data. However, to implement this approach, many hurdles must be overcome—ones that CEA-List has been working on for years. In 2025, we developed a virtual FHE processor, codenamed MadPanthera, that manipulates 8-bit integer ciphers, with results exceeding the current state of the art.



DAPHNÉ TRAMA,
PhD student



RENAUD SIRDEY
Research Director

CEA-List had been working on its open-source compiler, Cingulata, since 2015. In 2020, our researchers began looking for a way to give general-purpose FHE a “second wind”. The idea was to transform the promise of running any algorithm in the encrypted domain into a reality while seeking approaches that could exceed the state of the art in terms of potential hardware acceleration.

In 2022, Daphné Trama’s PhD research presented an opportunity to address these issues. The strategy was to create a small corpus of FHE operators based on the TFHE fully-homomorphic encryption scheme to rebuild, from the ground up, a more efficient, beyond-state-of-the-art general-purpose FHE approach. This meant creating a first set of general-purpose instructions for manipulating FHE ciphers for a software abstraction of an 8-bit processor.

We reached a major milestone this year with a solution based on advanced programmable bootstrapping techniques and the TFHE cryptosystem, which CEA-List helped develop in 2016-2017. One of the major challenges faced by the FHE community in recent years has been to gain a deeper understanding of TFHE’s scope and make the best use of its capabilities. To address these challenges, we

investigated the use of TFHE bootstrapping with integer ciphers—an approach that remains underexplored for general-purpose FHE and that is competitive with the state of the art. Like Cingulata, this method is based on the evaluation of Boolean circuits on binary ciphers.

Daphné’s team got to work developing a general-purpose homomorphic «processor» with an ISA (instruction set architecture) of around fifty instructions. This universal set of instructions was used to implement the homomorphic equivalent of any classical algorithm (including array sorting, averaging, finding a maximum/minimum, and for the first time, evaluating a neuron with a sigmoid activation function and inputs/outputs to several decimal places of precision). The ISA was used to evaluate a half-dozen algorithms. The results revealed that our approach is, on average, almost twice as efficient as the state of the art, performing particularly well in cases where the Boolean circuit corresponding to the executed instruction grows (typically, a multiplier of 1x is sufficient).

Cingulata, which remains significantly faster than competing solutions, is the most competitive with the state of the art.

The following table presents an overview of the results, with runtimes expressed in seconds.

Algorithms	MadPanthera (CEA-List, 2025)	Juliet	E3	Cingulata (CEA-List, 2020)	Concrete
Maximum	4,74 s	7,91 s	3,13 s	1,87 s	21,52 s
BubbleSort	15,22 s	31,62 s	12,70 s	6,18 s	78,07 s
SquaresSum	5,66 s	25,10 s	8,82 s	8,40 s	7,69 s
Average	2,36 s		3,57 s	10,31 s	
Sigmoid	8,98 s			969,0 s	



Daphné Trama presenting at CHES 2025, a leading international conference, in Kuala Lumpur.

It also provided an idea of what the computing kernels to be optimized and/or accelerated for general-purpose FHE might look like. These insights open the door to further joint research with CEA-List’s hardware design experts. In the longer term, this research could even lead to the development of a hardware accelerator specifically for the optimization of this virtual processor.


In connection with this research, we also addressed the homomorphic execution of the AES (Advanced Encryption Standard), a flagship use case at the center of the first NIST (National Institute of Standards and Technology) call for projects on FHE. After improving the international state of the art by a factor of ten in 2023^[3], several research teams from different countries have continued to come up with new techniques, setting a series of new «records» for the fastest homomorphic AES execution. In 2025, a partnership with CryptoExperts put France back in the lead in this friendly international competition with a new, more efficient technique that outperformed the previous record by 33%. The improvement was obtained by using MadPanthera’s foundations in conjunction with a complementary method called p-encoding.


50% Homomorphic computing algorithm runtimes on ciphers cut by half

• • ➤ **Major project and/or partnership**
The research was part of the national PEPR SecureCompute and TRUSTIncloudS projects.

• • ➤ **Flagship publications**
This research was presented at CHES 2025, a leading conference, and was published [1] in the conference proceedings (IACR Transactions on Cryptographic Hardware and Embedded Systems).

 [1] «**Designing a General-Purpose 8-bit (T)FHE Processor Abstraction**». Daphné Trama, Aymen Boudguiga, Pierre-Emmanuel Clet, Renaud Sirdey, Nicolas Ye: *IACR Trans. Cryptogr. Hardw. Embed. Syst.* 2025(2): 535-578 (2025) <https://doi.org/10.46586/tches.v2025.i2.535-578>

 [2] «**Further Improvements in AES Execution over TFHE**». Sonia Belaïd, Nicolas Bon, Aymen Boudguiga, Renaud Sirdey, Daphné Trama, Nicolas Ye: *IACR Commun. Cryptol.* 2(1): 39 (2025) <https://doi.org/10.62056/ahmp-4tw9>

 [3] «**A Homomorphic AES Evaluation in Less than 30 Seconds by Means of TFHE**». Daphné Trama, Pierre-Emmanuel Clet, Aymen Boudguiga, Renaud Sirdey: *WAHC@CCS 2023: 79-90* <https://doi.org/10.1145/3605759.3625260>

Cybersecurity and trustworthy digital technology

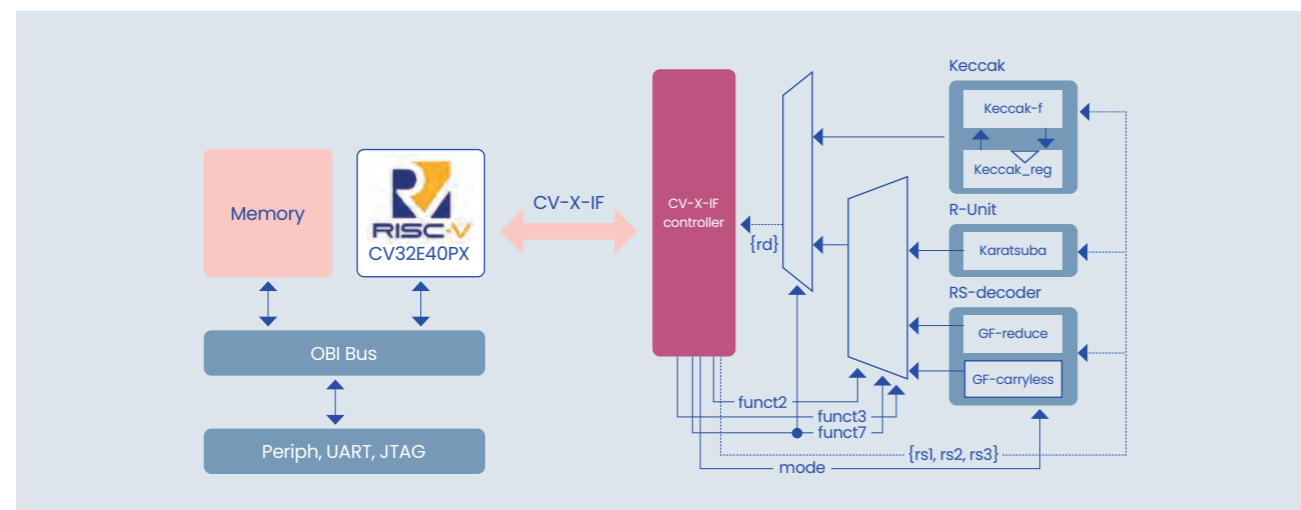
Security-enabling technologies

Hardware accelerator tightly coupled to a RISC-V core for post-quantum cryptography

With the development of quantum computers—and their threat to modern cryptography—post-quantum cryptography is gradually making inroads into the industrial ecosystem. HQC (Hamming Quasi-Cyclic) is one of the post-quantum encryption algorithms recently standardized by NIST. The algorithm, based on a mathematical formalism derived from code theory, exhibits relatively high runtimes when deployed on the microcontrollers found in embedded systems. We developed TYRCA, an HQC accelerator designed specifically for RISC-V processors. A CV-X-IF interface is used to tightly couple the accelerator to the processor core. TYRCA sped up the algorithm’s runtime by 94% to 95 % and, on an FPGA implementation, obtained a beyond-state-of-the-art performance-to-area ratio.



IN PARTNERSHIP WITH
CEA-LETI AND
POLITECNICO DI TORINO



Architecture diagram of a system-on-chip with the TYRCA coprocessor coupled to the RISC-V CV32E40PX core. The coprocessor consists of a CV-X-IF controller and set of hardware accelerators for specific HQC operations, including Keccak, polynomial multiplication (Karatsuba), and Reed-Solomon (RS) decoding.

Quantum computers, which have the potential to revolutionize medicine, materials science, and artificial intelligence, also represent a threat to the security of our communications. Cryptographic systems like RSA, Diffie-Hellman, and ECC have been shown to be vulnerable to quantum attacks. Post-quantum cryptography (PQC) aims to develop systems resistant to these attacks.

The HQC algorithm, which is based on error-correction codes, was selected as the standard by NIST in 2025. The proposed methods for HQC software implementation perform relatively poorly compared to alternative Euclidean-network-based solutions. Hardware implementations, while faster, rely on polynomial multiplication involving very large polynomials requiring a very large silicon area—the algorithm’s main bottleneck.

Tightly-coupled acceleration, which encapsulates the algorithm’s recurring operations in custom instructions added to the processor instruction set, offers an alternative to conventional hardware acceleration methods. The accelerator can be tightly integrated into the RISC-V processor execution pipeline using the Core-V eXtension Interface (CV-X-IF). This solution offers transparent speed increases while overcoming the traditional hurdles to ISA extension.

We applied this new, tightly-coupled acceleration strategy and designed three dedicated hardware acceleration technologies targeting the HQC algorithm’s main bottlenecks. The first, R-Unit, uses a multi-level Karatsuba algorithm to speed up polynomial multiplication on 32-bit blocks with four custom instructions to ensure that the results are completely controlled. The second, RS-Decoder, contains several specialized instructions to eliminate loops and intermediate results, speeding up key Galois field operations like carry-free multiplication and final zero counting. The third, the Keccak accelerator, leverages a dedicated register and three custom instructions to ensure efficient loading, processing, and storage. This minimizes overhead compared to conventional loosely-coupled approaches, for effective management of the 1,600-bit permutation state.

TYRCA, with its tightly-coupled acceleration, delivers substantial performance improvements over the original HQC software implementation. The number of clock cycles in key generation (KeyGen) and encapsulation and decapsulation (Encaps, Decaps) operations is reduced by around 95% at all security levels (HQC-128/192/256). This approach also substantially reduces instruction memory use. Implemented on a Kintex-7 FPGA target, TYRCA occupies less than 26% of the total system-on-chip area. R-Unit, which delivers the highest performance improvement, takes up less than 10 % of the area. Normalized velocity metrics (velocity/surface) confirm that TYRCA outperforms the existing loosely-coupled approaches.



Flagship publication

«TYRCA: A RISC-V Tightly-coupled accelerator for Code-based Cryptography». Dolmeta, A., Di Matteo, S., Valea, E., Carmona, M., Loiseau, A., Martina, M., & Maserà, G. (2025, March). In *2025 Design, Automation & Test in Europe Conference (DATE)* (pp. 1-7). <https://doi.org/10.23919/DATE64628.2025.10993202>
This paper was selected as Best Paper Candidate at the DATE conference (12 papers selected out of 330 accepted at the conference and over 1,200 submissions). The results published show speeds 20 times faster than the reference software implementation of the HQC post-quantum cryptography algorithm.

Figure 1: RISC-V pipeline integration strategy with CV-X-IF

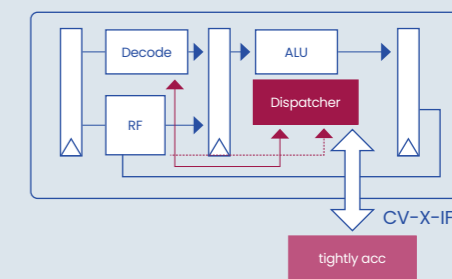


Figure 2: Multi-level Karatsuba decomposition in R-unit

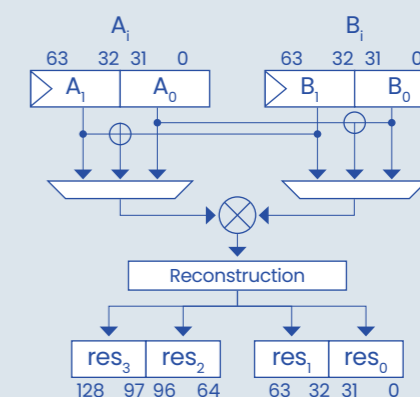


Figure 3: Reduction in total number of cycles on HQC showing performance gains

Version	Level	KGen+Enc+Dec
SW	128	407.90M
SW	192	1188.77M
SW	256	2182.38M
TYRCA	128	20.50M [-95%]
TYRCA	192	68.32M [-94%]
TYRCA	256	126.61M [-94%]

Come to **List Days** **Conference 2026 !** and discover all our research results!

PROGRAM

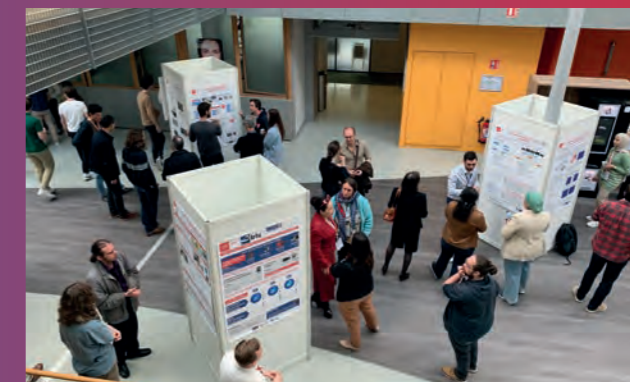
Introductory keynotes

Workshops on a variety of topics:

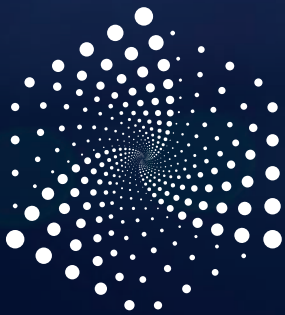
- Sensors and instrumentation for physical measurement
- HPC & quantum computing
- Data and systems engineering
- Smart systems and cooperation with humans
- Toward more realistic AI
- Eco-innovation
- Cybersecurity
- AI for embedded systems
- Toward more reliable use of AI
- Physical modeling and simulation of critical systems and components



Two interactive poster
sessions presenting
our PhD students' research



June 10 & 11, 2026 – Nano-INNOV – Palaiseau
IMMERSE YOURSELF IN SCIENCE



Scientific and technological
excellence



Building and leading ecosystems
at the national and European levels



Responding to the needs
of industry



list.cea.fr/en/