



Simplifying AI at the Edge.
Accelerating Computer Vision.

Ubiquitous AI Computing - AI Everywhere

Giuseppe Garcea, PhD, Director of silicon and Member fo the founding team

19 Nov 2025

Leading AI computing at the edge and beyond

Company

Founded by a team from IBM, IMEC, ETH Zurich and Bitfury AI, Axelera AI is the largest AI semiconductor startup in Europe

2021

Founded in July

\$125M

Funding to date

\$60M

EU Grants

200+

Employees

60+

PhDs

15

Countries

Product

We build the most powerful, efficient and affordable AI Processing Unit for computer vision & GenAI



METIS

First Gen (2024)

5x

More power
efficient

2-5x

More cost
efficient

1-3x

Higher
performance



EUROPA
Generative AI
(2025/6)



TITANIA
EuroBEC
(2027/8)

Market

We have validated product-market-fit and are executing against a strong go-to-market strategy



€40B+

Market opportunity



€300M+

Total pipeline



150+

Customers

Product
execution



Three chips
taped out in
three years

Intellectual
property

10+

Strategic Pending
Patents

Capital
efficiency

\$80M

From idea to
commercialization

Our investors

Innovation Industries

INVESTNL

**SAMSUNG
CATALYST
FUND**

imec xpand

**VERVE
VENTURES**

BITFURY

cdp"

Fondo Nazionale Innovazione
CDP Venture Capital Sgr

European
Innovation
Council



sfpi fpim

Our team

IBM

Lenovo

imec

Movidius

NVIDIA

intel

Google

Ambarella

ETH zürich

arm

DELL

blaze

GRAPHCORE

BCG

AXELERA
ARTIFICIAL INTELLIGENCE



Retail



Security



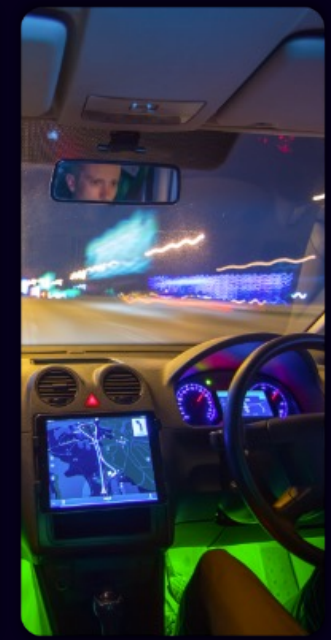
Robotics
& Drones



Industry 4.0



Healthcare



Automotive

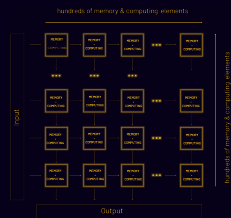
A powerful, efficient, cost-effective and user-friendly **hardware and software**
AI platform to accelerate **computer vision and generative AI** in edge devices

Technology

Digital in-memory computing and RISC-V deliver high performance and programmability

Digital In-Memory Computing

Densely interleaved memory and compute elements reduce data movement, physical size, and increase throughput for MVM



Matrix multiply

Massively parallel execution
= **High Throughput**

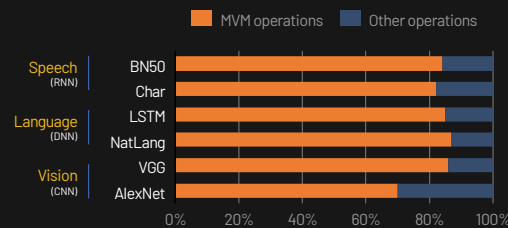
Weight cells and compute merge
= **Small Die Area**

Local data & low data movement
= **Power-efficient**

INT8 weights/activations and INT32 accumulation
~ **FP32 Level Accuracy**

Neural Network Workloads

Matrix vector multiplications (MVM) constitute 70-90% of neural network operations.



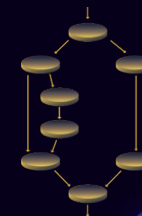
Source: B. Fleischer, VLSI/19

So how do we balance performance, programming efficiency and silicon cost?

RISC-V Control

Custom designed RISC-V control system is responsible for general purpose operations, scheduling of memory and graph processing

Data flow control



RISC-V®

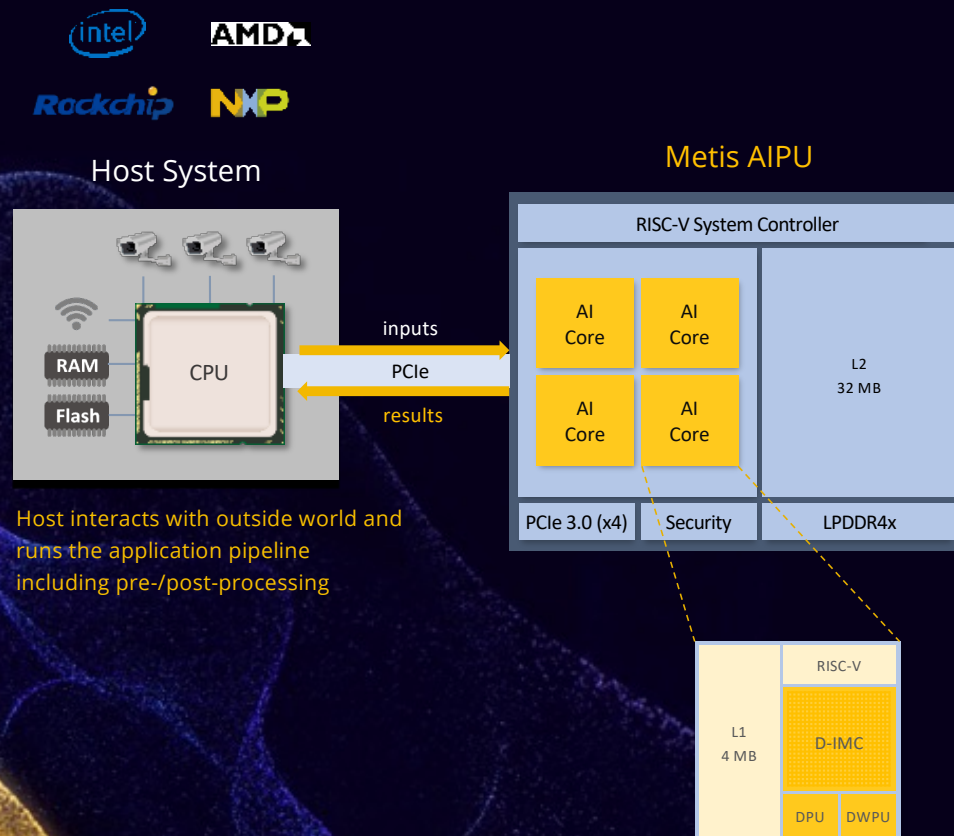
Vector processing extensions
= **Future-proof**

Data prefetching
= **High Performance**

Highly programmable
= **Usability**

Open source
= **Collaborative improvement/innovation**

Metis AIPU – system block diagram



- **AI Engine with four AI Cores**
 - D-IMC : matrix-vector multiply, densely interleaved memory and compute with >1M elements
 - DPU : element-wise vector operations and activations
 - DWPU : sparse ops; pooling, depth-wise convolution, up-sampling etc.
 - AI Cores run independently or in concert
- **Hierarchical compiler-managed memory**
 - Execute-overlapping prefetch and data reuse allocation
 - D-IMC : 4 x 1 MB
 - L1 AI Cores : 4 x 4MB
 - L2 Shared : 32 MB
 - DDR : Up to 4 GB to fit large models, minimize latency and context switching overhead
- **Security unit**
 - Secure key storage and management, encryption and decryption
- **Performance and physical data**
 - 214 tera-operations-per-second (TOPS)
 - 4-7W typical power consumption
 - 12 nm CMOS technology
 - -40°C to +85°C extended temperature range

Sample Pipelines & Use cases

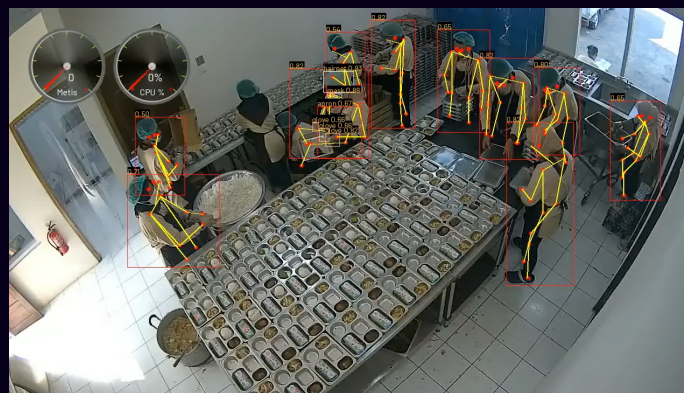
- Parallel and cascaded pipelines can be built via the simplified user journey
- Support for additional use cases is increasingly **generalized**

```
pipeline:
  - master_detections:
    - model_name: yolov8spose-coco-onnx
    - input:
      - type: image
    - preprocess:
      - letterbox:
      - torch-totensor:
    - postprocess:
      - decodeyolopose:
        - conf_threshold: 0.25
        - nms_iou_threshold: 0.45
    - secondary_detections:
      - model_name: yolov8n-weapons-and-knives
      - input:
        - type: image
        - source: roi
        - where: master_detections
        - which: AREA
        - top_k: 10
        - expand_margin: 0
      - preprocess:
        - letterbox:
        - torch-totensor:
      - postprocess:
        - decodeyolo:
          - conf_threshold: 0.25
          - nms_iou_threshold: 0.45
```

Pipeline for gun detection



Cascaded pipeline – pose estimation followed by detection in zoomed-in region of interest (ROI)
Applied to different use cases (fruit detection, gun detection)



Pipeline for lead customer



Parallel pipeline example:
pose estimation and face detection in parallel

Small Language Models running on Metis



Tell me everything you know about Axelera AI



Axelera AI is a forward-thinking technology company that specializes in the development of advanced AI solutions. Their flagship product, the Metis chip, is a state-of-the-art AI accelerator designed to significantly enhance the processing power and efficiency of AI applications.

- First demonstration of a generative workload on Metis hardware
- This milestone is a first step towards achieving 3 key company goals:
 1. Market Metis's genAI Capabilities towards customers and **investors, e.g. at CES'25 together with Arduino**
 2. Verify LLM performance on **Europa** prior to tape-out (10x perf)
 3. Understand bottlenecks of the current systems, towards the architectural definition of Titania (100x perf)



Product Lineup

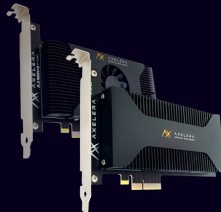
Modules



Metis M.2

AI acceleration to systems with an M.2 2280M slot where space is at a premium

Cards



Metis PCIe

PCIe cards with 1x or 4x Metis AIPUs for Edge Servers where AI performance and flexibility is a priority

Boards



Single Board Computer

ARM (Rockchip RK3588)

For stand-alone and compact form factor embedded systems

Systems



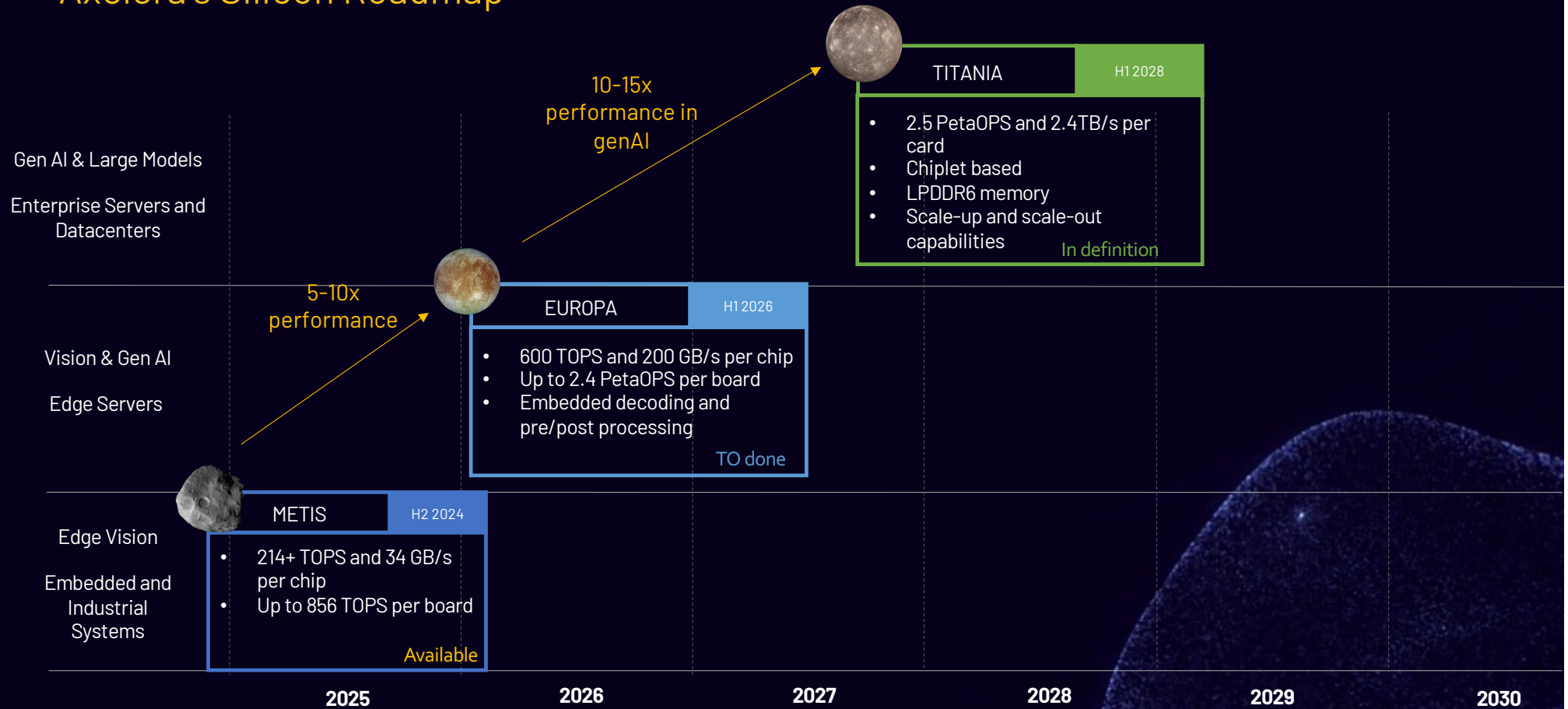
Partner products

x86 Edge Servers, Industrial PC's

Ready to use devices for edge or near edge processing where out-of-the-box systems are needed

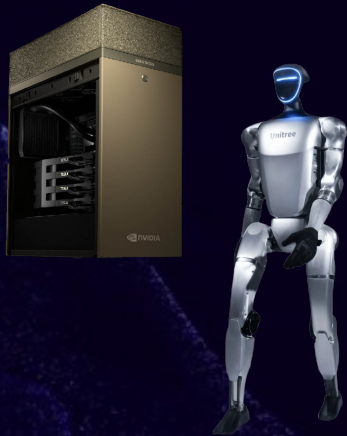


Axelera's Silicon Roadmap



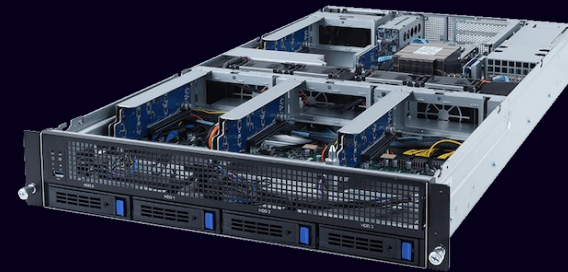
Europa: two markets

Mid-end Edge



Desktops, workstations and Robotics requiring high performance at low power and reasonable cost

Enterprise AI Server



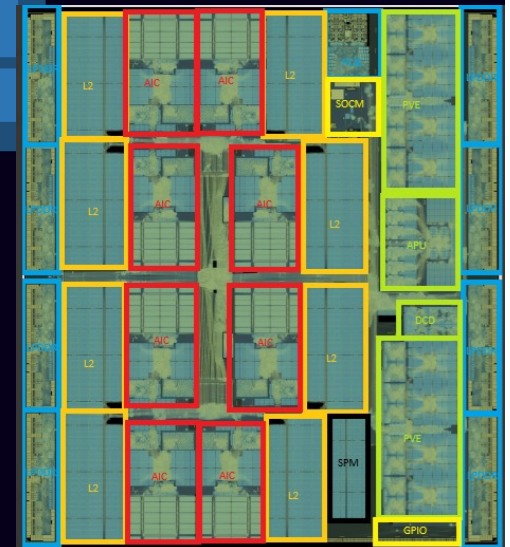
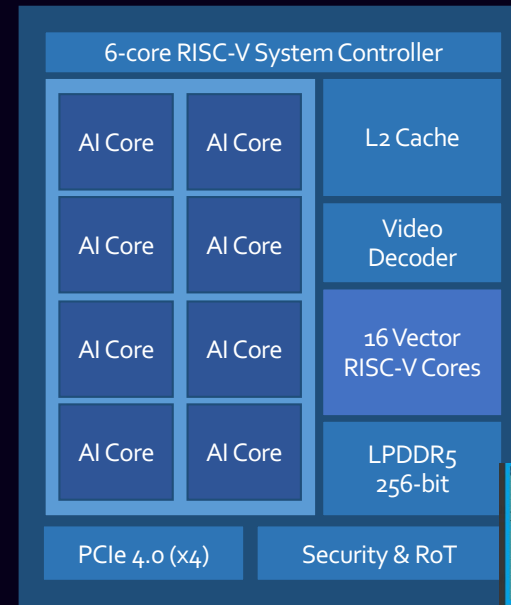
Enterprises that want to own their own AI inference infrastructure but need it at a sane price point.



Europa AIPU

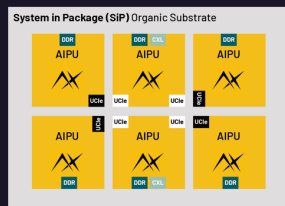
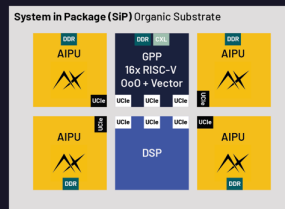
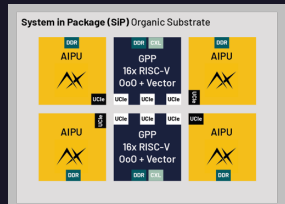
Block Diagram And Characteristics

Peak AI performance	629 TOPs @ INT8 (1.2 GHz) 314 TOPs @ INT16 (A) / INT8 (W)
Internal memory (AI and Vector Cores)	48 MB L1 SRAM 128 MB L2 SRAM
External memory	LPDDR5, 200GB/s
Communication bus	PCIe 4x Gen4
Security module	Root of Trust (RoT)
Video decoder	HEVC265 Decoder, 4K120FPS
Pre/post-processing	16x RISC-V Vector Processors
System controller	Quad-Core RISC-V cluster



- ✓ **High performance for both edge and cloud applications:** 4-5x more compute power, 5.5x more memory BW
- ✓ **Lower Total Cost of ownership:** integrated video decoder and pre-/post-processing capabilities
- ✓ **Future proof:** INT16 type support, improved transformer support, vector compute capabilities

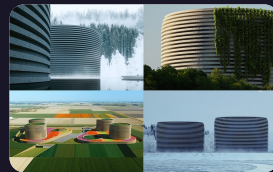
Axelera AI – target market and applications for TITANIA



HPC



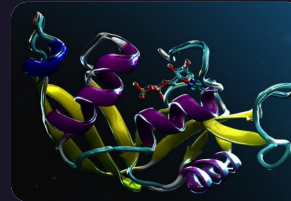
Data Center



Enterprise Servers



Drug Discovery



Agri & Climatech



Scientific Research



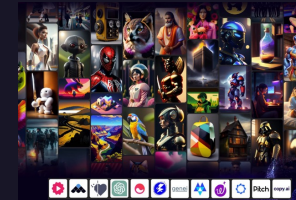
Finance



Cloud Gaming



Generative AI

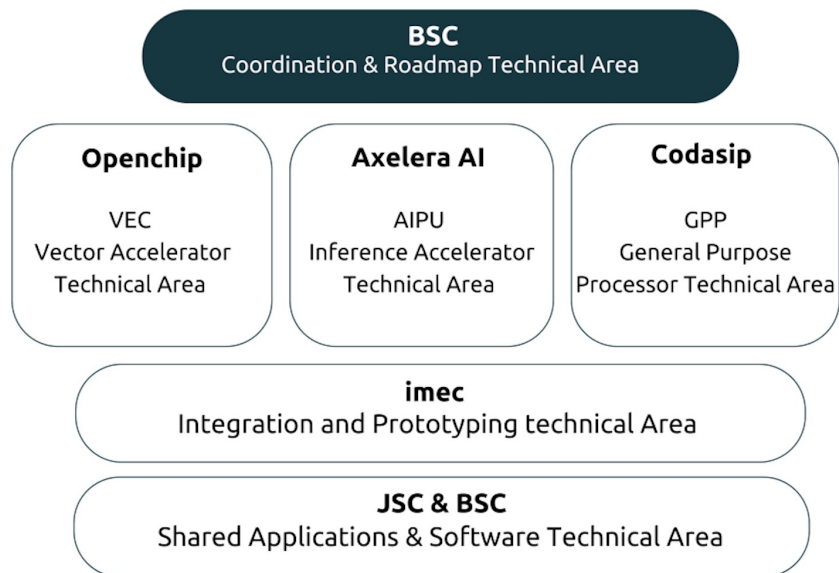


AMD CEO Lisa Su believes the AI semiconductor industry could reach [\\$400bn by 2027](#)

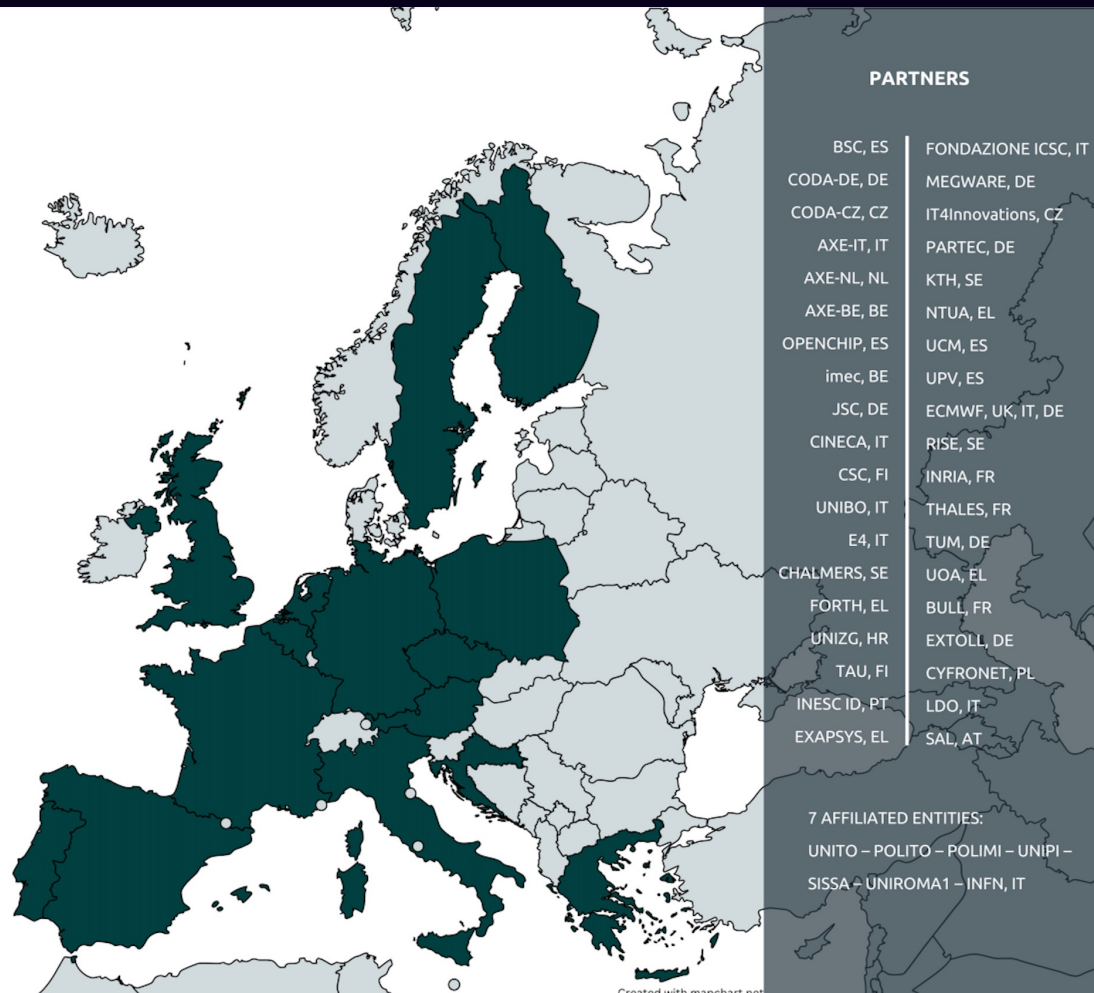
Titania Part of JU EruoHPC Dare project



DARE SGA1 Project Structure - Technical Areas



The Digital Autonomy with RISC-V in Europe (DARE) project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 101202459. The JU receives support from the European Union's Horizon Europe research and innovation programme and Spain, Germany, Czechia, Italy, Netherlands, Belgium, Finland, Greece, Croatia, Portugal, Poland, Sweden, France and Austria.





Thank You!