



# HiPEAC Vision 2024

HIGH PERFORMANCE, EDGE AND CLOUD COMPUTING

September 5<sup>th</sup>,  
2024

Marc Duranton (HiPEAC Vision coordinator)



The **HiPEAC Vision** is a deliverable of the **coordination and support action on High Performance, Edge And Cloud** computing (previously: **High Performance and Embedded Architecture and Compilation**), a European network of almost 2,000 world-class computing systems researchers, industry representatives and students.

The HiPEAC Vision shows the trends, technology evolutions and limitations and position of Europe in the domain of computing (hardware and software) and provides recommendations to the HiPEAC community at large.



January 2024 version is available at:

<https://www.hipeac.net/vision/>



# HiPEAC Vision 2024 overview

A sense of urgency: technology is evolving faster than humans' natural pace, in a context of intensifying geopolitical and environmental pressures.

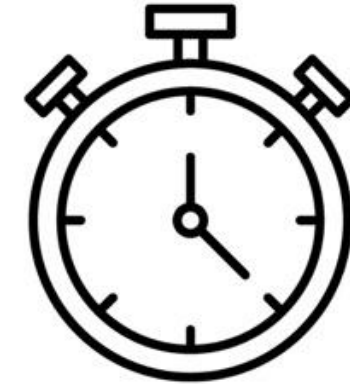
We are in races:

against **time**

- because of external constraints – e.g. global warming
- because the evolution of technology is very fast (e.g. AI)

with **the rest of the world**

- on multiple aspects (economy, reducing dependances, ethics, ...)



The HiPEAC Vision identified specific races:

- **Race for the “next web”, the “next computing paradigm”** – the continuum of computing
- Race for **artificial intelligence**
- Race for innovative and new **hardware** (including quantum)
- Race for **cybersecurity**
- Race for **sustainability**
- Race for **sovereignty** ⇨ So important that it is embedded in each other section

And a set of global recommendations promoting **holistic approaches and breaking silos**



---

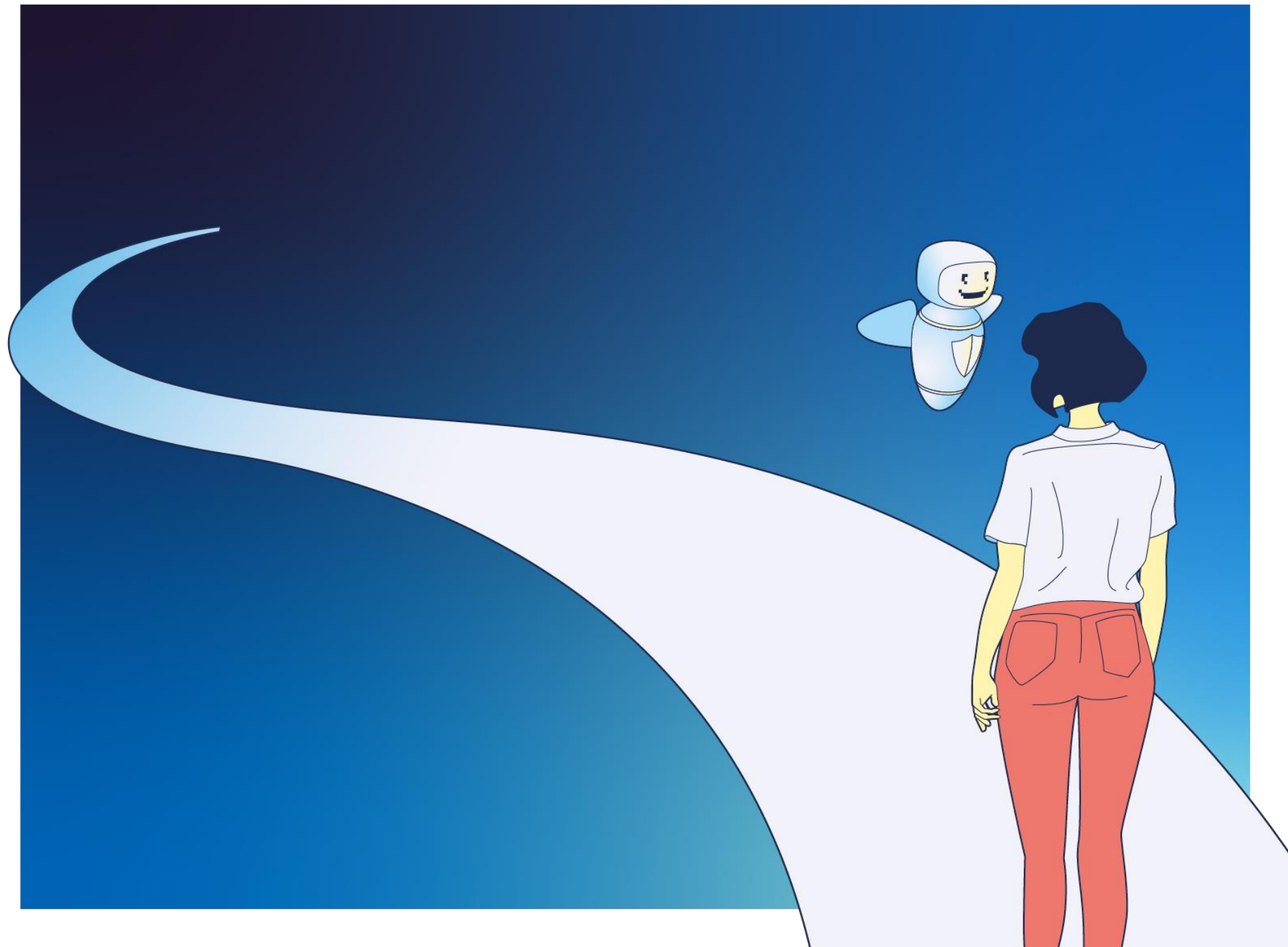
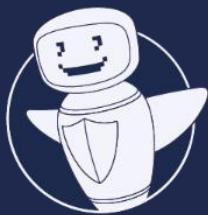
## HiPEAC Vision 2024

HIGH PERFORMANCE,  
EDGE AND CLOUD COMPUTING

---

**NEXT  
COMPUTING  
PARADIGM**

---





## HiPEAC Vision 2024

HIGH PERFORMANCE,  
EDGE AND CLOUD COMPUTING

## NEXT COMPUTING PARADIGM

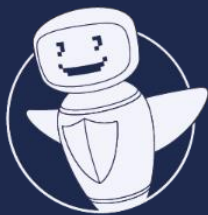




## HiPEAC Vision 2024

HIGH PERFORMANCE,  
EDGE AND CLOUD COMPUTING

## NEXT COMPUTING PARADIGM



# The next computing paradigm...

After

- the **Web**,
- CyberPhysical Systems (**CPS**),
- the **Cloud**,
- the Internet of Things (**IoT**),
- **Digital Twins**,
- the **continuum of computing**,
- the **metaverse**,
- Artificial Intelligence (**AI**)...

HiPEAC conjectures that the next evolution will be a **convergence** of the key elements of all previous technologies into



Image created by Dall-E 3

## The next computing paradigm – NCP

Or

*“How computing will experiment the world”*

# Key observations for the “next computing paradigm”: NCP

## Data & interconnected machines

- Increasing amounts of data are being generated by interconnected machines to create new services. The “NCP” will have to integrate the “web of machines” with the “web of humans”.

## Large scale safety-critical systems and digital twins

- Systems will be increasingly intertwined with the physical world. Real-world constraints such as safety, real time and location will need to be considered. Digital twins - to model past, present and future) will be key for improving efficiency.

## The spatial web

- Systems will also be spatial: we will go/are into the computing continuum (ubiquitous computing). Properties depend on location and time (4D).

## Everything as a service

- Key evolution: “everything as a service” (XaaS). Applications will be composed by tailored coordination of services (including digital twins) that will run where it is the most efficient according to user’s criteria.

## Smart orchestrators

- Smart orchestrators will be required to orchestrate services while protecting and guiding users into the complexity of services in a trustable way. This will also use AI techniques for smooth and natural communication with humans (e.g. LLMs and LAMs).



In a circular and sustainable economy



# What are the key Ingredients? - 1

- Applications are defined as (dynamic) composition of various services (Xaas)
- The services are selected and orchestrated together
- The orchestration is done in several steps:
  - Decomposition of the application and analysis of the required services
    - Can be explicitly done – software (e.g. Python calling libraries through API = services)
    - Or automatically – by AI – LLM code generation feature
  - Discovery of available services (distributed, through 3<sup>rd</sup> party, ....)
  - Selection of service according to defined set of criteria
  - Secure activation and authentication of the services, with guaranties
  - Orchestration of the services to perform the requested application



# What are the key Ingredients? - 2

- Independence of the services (software) and hardware
  - Code (*not only data*) can migrate from device to device
  - Hardware (processing, acceleration, storage) can be shared among devices, based on properties (processing power, QoS, energy efficiency, etc.)
- The orchestration/mapping service/resource to device is key
- This mechanism can be seen at multiple scales
  - Various functions (executed locally on a computer)
  - Specific accelerators as chiplets orchestrated by a processor in a chiplet+interposer solution
  - Functions in a device (smartphone/PC) (like camera, display, processing, storage, security, network)
    - Authentication also at this level, e.g. Apple devices
  - Resources sharing between devices (*“devices are aggregated at the super device, allowing flexible scaling of device hardware capabilities”*)
  - Resource sharing at company/premise level
  - Federated cloud
  - Aggregation of data centers (e.g. like for Destination program)
- **Edge to cloud continuity “continuum”**

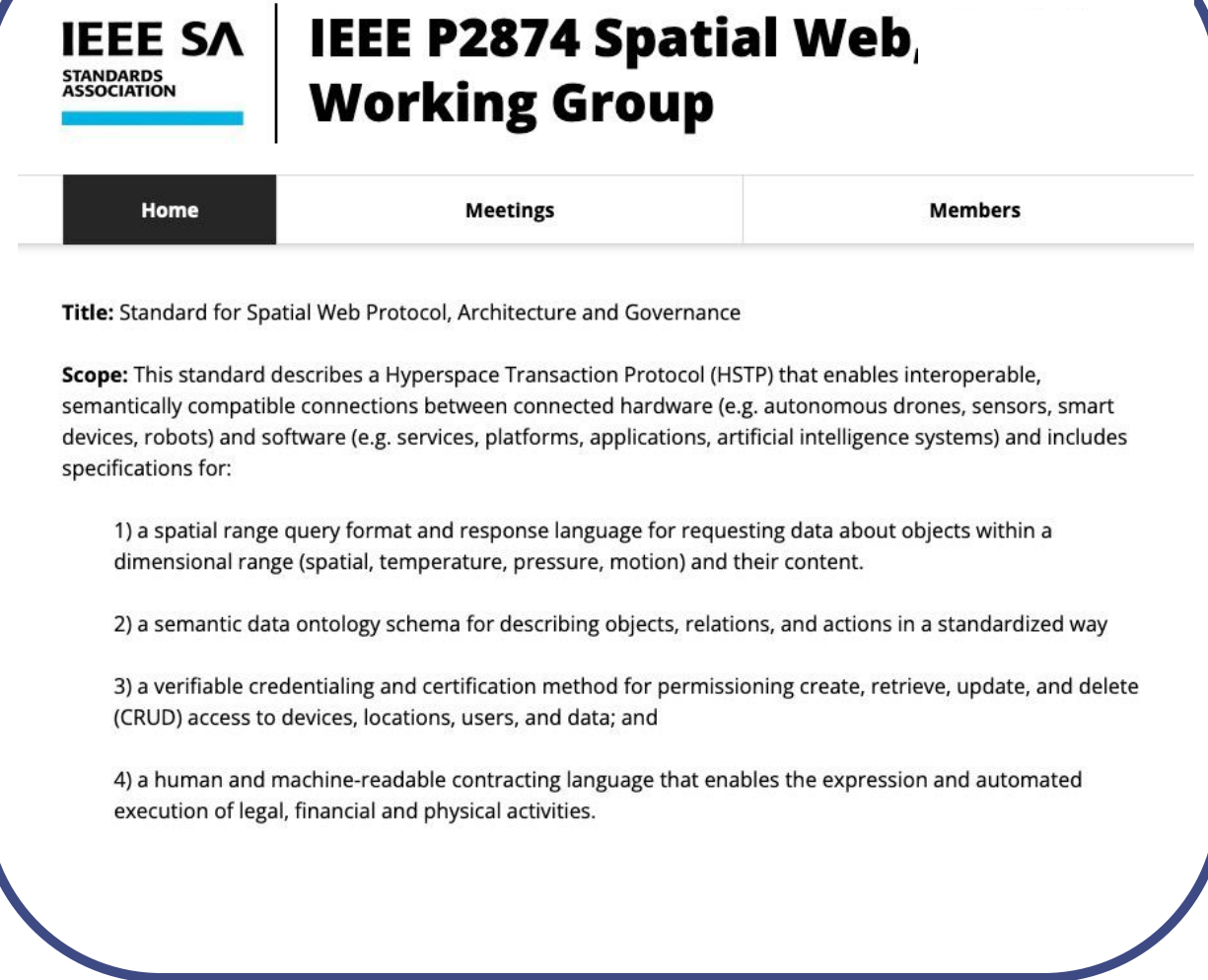
*“users can integrate capabilities of their various smart devices, implementing ultra-fast connection, capability collaboration, and resource sharing among them. This way, services can be seamlessly transferred to the most suitable device, delivering smooth all-scenario experience.”*

From harmonyos-guides-v3



# What are the key Ingredients? - 3

- Non functional properties are key elements to select the service
  - Latency (interaction with the physical world, e.g.
  - Performance
  - Privacy, locality
  - Cost
  - Security
  - Governance/mediation
  - Trust
  - Eco criteria
  - ...
- Current protocols should be extended to support
  - E.g. IEEE P2874 for spatial web
- They should be hardware enforced
  - Virtualization
  - Silo, trust zones
  - Cryptographic protocols
  - (Lightweight) authentication
  - ...



**IEEE SA**  
STANDARDS  
ASSOCIATION

## IEEE P2874 Spatial Web, Working Group

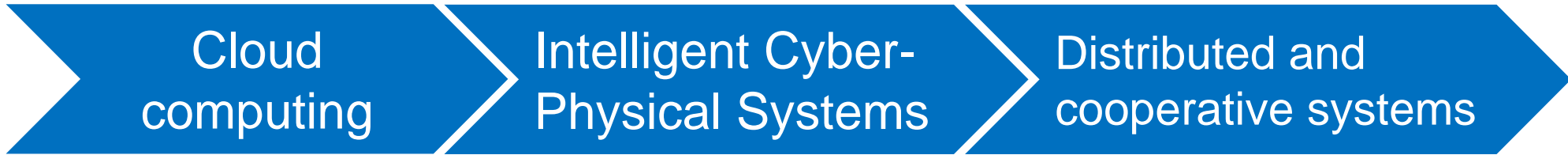
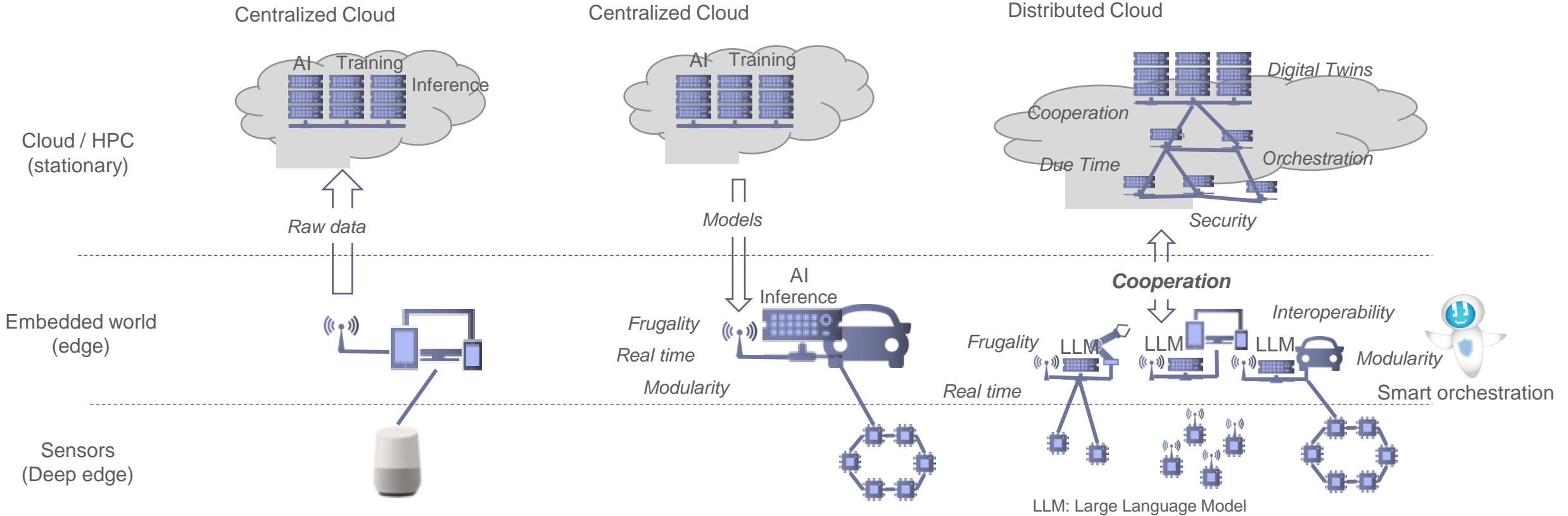
**Home** Meetings Members

**Title:** Standard for Spatial Web Protocol, Architecture and Governance

**Scope:** This standard describes a Hyperspace Transaction Protocol (HSTP) that enables interoperable, semantically compatible connections between connected hardware (e.g. autonomous drones, sensors, smart devices, robots) and software (e.g. services, platforms, applications, artificial intelligence systems) and includes specifications for:

- 1) a spatial range query format and response language for requesting data about objects within a dimensional range (spatial, temperature, pressure, motion) and their content.
- 2) a semantic data ontology schema for describing objects, relations, and actions in a standardized way
- 3) a verifiable credentialing and certification method for permissioning create, retrieve, update, and delete (CRUD) access to devices, locations, users, and data; and
- 4) a human and machine-readable contracting language that enables the expression and automated execution of legal, financial and physical activities.

# Evolution of computing: Cloud, CPS, IoT, AI → Next Computing Paradigm





---

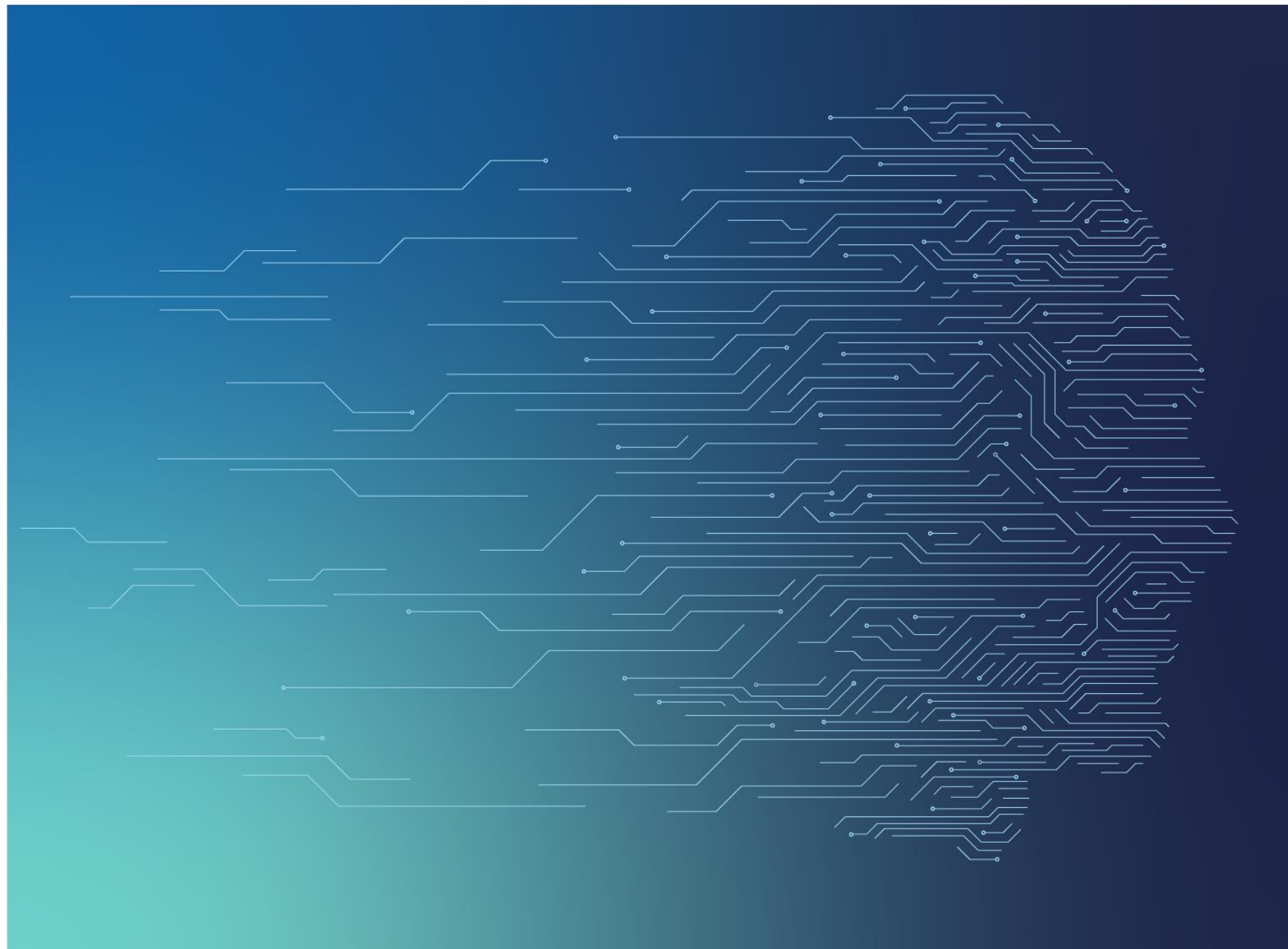
## HiPEAC Vision 2024

HIGH PERFORMANCE,  
EDGE AND CLOUD COMPUTING

---

## ARTIFICIAL INTELLIGENCE

---





---

## HiPEAC Vision 2024

HIGH PERFORMANCE,  
EDGE AND CLOUD COMPUTING

---

## ARTIFICIAL INTELLIGENCE

---





---

## HiPEAC Vision 2024

HIGH PERFORMANCE,  
EDGE AND CLOUD COMPUTING

---

## ARTIFICIAL INTELLIGENCE

---



# Recommendations for artificial intelligence

- ❑ LLMs applications are growing at a rapid pace, and will cover more and more applications domains (multimodality)
  - ❑ **Europe should be an active player, by providing foundation models** and ways to fine tune models into **specialized agents (Agentic AI)**
  - ❑ **Open source** is important (e.g. Hugging Face)
- ❑ (few G parameters) **LLMs are already running at the edge**, special accelerators will be here. Europe should develop its own solutions!
- ❑ These **Specialized Agent Models (SAM) should be orchestrated using similar approaches that in the NCP**
- ❑ LLMs will help (also hardware, to be coming) software development, in two directions:
  - ❑ We have a huge problem to find specialists in HiPEAC domains, so **AI could improve productivity** of existing engineers/researchers, but they should embrace it with training to use AI efficiently
  - ❑ **Making the digital world accessible for everybody** (e.g., by orchestrating APIs, generating Python Code,...).
    - ❑ This a societal issue and should be available for every European citizen.
    - ❑ "Centaur" for developers, improving productivity, better quality of code
- ❑ Develop methods to ensure "automagically" that the generated code, HW is "**correct by construction**", valid etc without human checking?



# And we should care of the societal impacts...





---

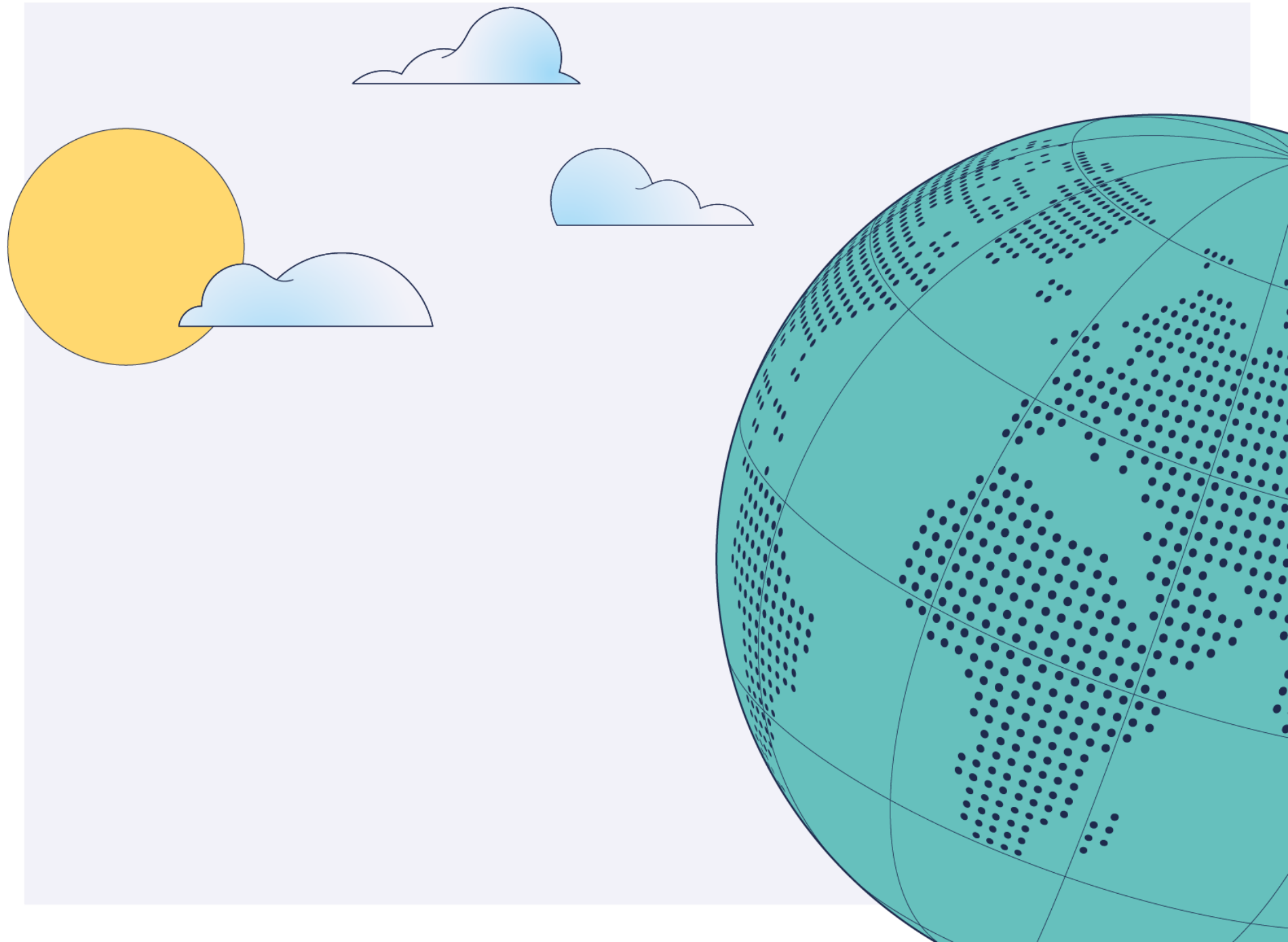
## HiPEAC Vision 2024

HIGH PERFORMANCE,  
EDGE AND CLOUD COMPUTING

---

## SUSTAINABILITY

---





---

## HiPEAC Vision 2024

HIGH PERFORMANCE,  
EDGE AND CLOUD COMPUTING

---

## SUSTAINABILITY

---





## HiPEAC Vision 2024

HIGH PERFORMANCE,  
EDGE AND CLOUD COMPUTING

## SUSTAINABILITY

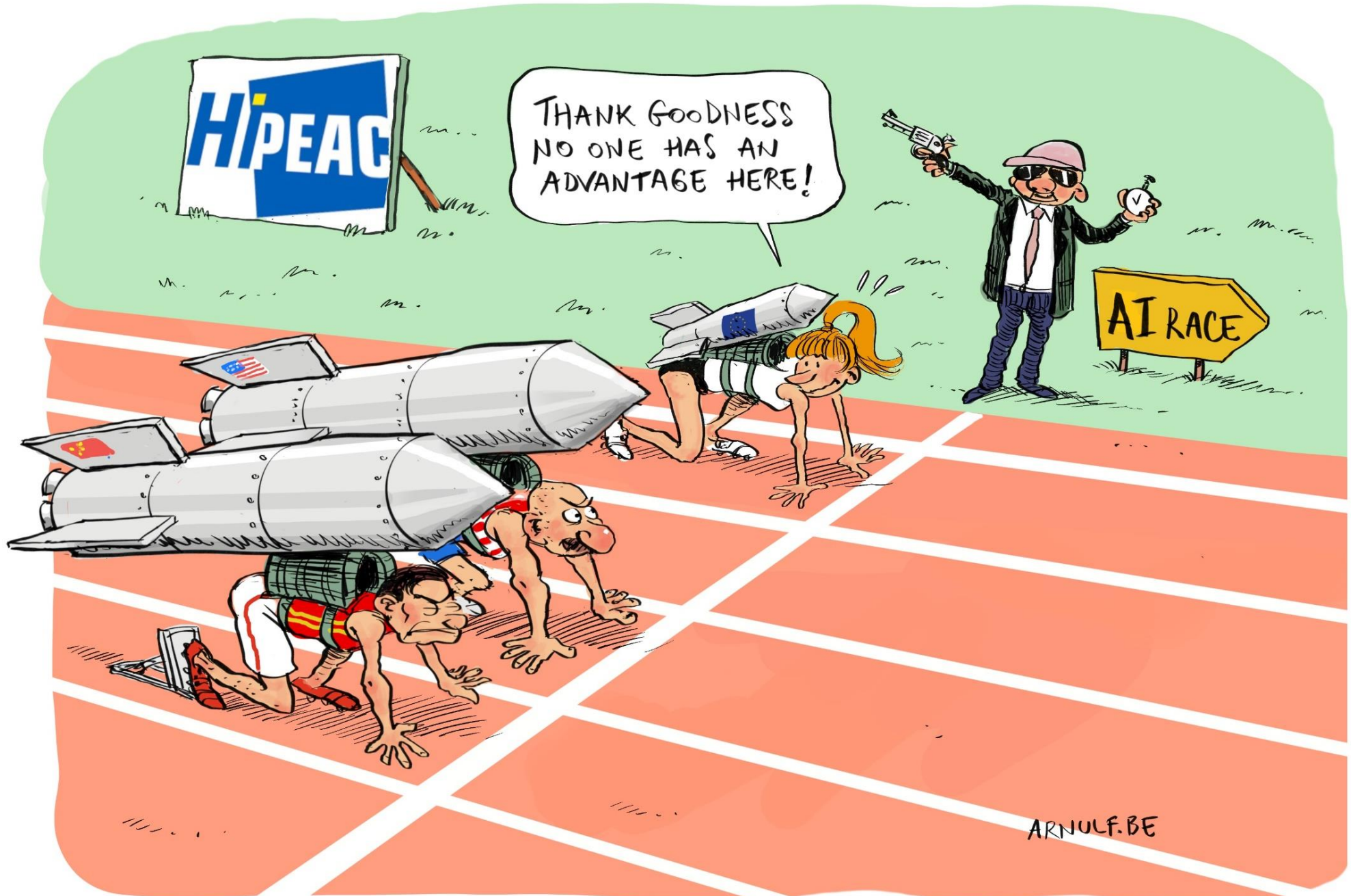


# The race for sustainability

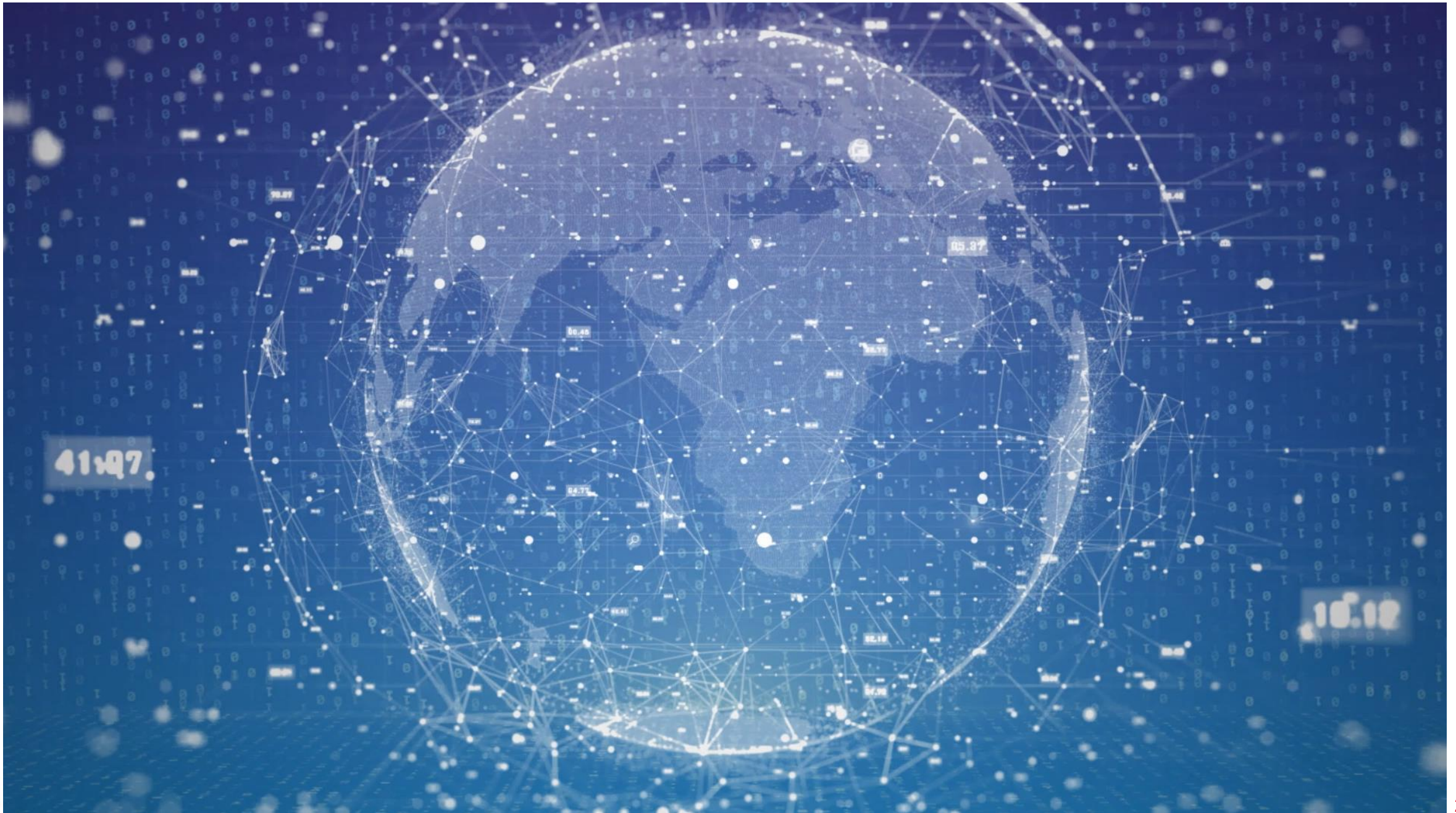


...the challenge

# CONCLUSION: WE LIVE A CHALLENGING TIME!



# HiPEAC Video





# HiPEAC Vision 2024

HIGH PERFORMANCE, EDGE AND CLOUD COMPUTING

Full version:

[hipeac.net/vision](https://hipeac.net/vision)

Articles on specific topics:

<https://vision.hipeac.net/>