



**HAL**  
open science

## HiPEAC Vision 2024

Marc Duranton, Paul Carpenter, Koen de Bosschere, Thomas Hoberg, Charles Robinson, Tullio Vardanega, Olivier Zendra

► **To cite this version:**

Marc Duranton, Paul Carpenter, Koen de Bosschere, Thomas Hoberg, Charles Robinson, et al..  
HiPEAC Vision 2024. HiPEAC. , pp.24, 2024, 9789078427056. hal-04884248

**HAL Id: hal-04884248**

<https://inria.hal.science/hal-04884248v1>

Submitted on 13 Jan 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

# HiPEAC Vision 2024

HIGH PERFORMANCE, EDGE AND CLOUD COMPUTING



**Editorial board:**

Marc Duranton (editor in chief), Paul Carpenter, Koen De Bosschere, Thomas Hoberg, Charles Robinson, Tullio Vardanega, Olivier Zendra.



Funded by  
the European Union

This document was produced as a deliverable of the Horizon Europe HiPEAC ("High Performance, Edge And Cloud computing") CSA under grant agreement 101069836.

The editorial board is indebted to Dr Max Lemke and Jan Komarek of the Directorate-General for Communication Networks Content and Technology of the European Commission for their active support of this work.

Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those either of the full HiPEAC community or of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

The large language model ChatGPT was used to summarize and enhance some of the text in this document.

Design: [www.magelaan.be](http://www.magelaan.be) / Cover design: Roger Castro, Monzón Studio / Cartoons: Arnout Fierens, Arnulf.be  
© 2024 HiPEAC

ISBN 9789078427056

# Foreword

For the seventh edition of the European-funded HiPEAC project, the source of the acronym 'HiPEAC' changed from 'High Performance Embedded Architecture and Compilation' to 'High Performance, Edge And Cloud computing' to better reflect how computing has morphed from discrete operations into a 'continuum of computing'. As HiPEAC has evolved, so has the HiPEAC Vision: with many domains covered by the HiPEAC community changing rapidly, the HiPEAC Vision is now issued every year, rather than every two years.

This release, the tenth to date, builds on the core ideas developed in the HiPEAC Vision 2023. In particular, it provides an in-depth exploration of the 'next computing paradigm' (NCP)<sup>1</sup>, which further develops



the 'next web' concept of the HiPEAC Vision 2023. This 'next computing paradigm' merges aspects of the web, the cloud, cyber-physical systems, the internet of things, digital twins, the metaverse and artificial intelligence (AI) into a coherent continuum of computing, intertwining the real world with the cyberworld.

The core themes of the HiPEAC Vision 2024 centre around the dynamic and competitive landscape of technological progress, with a specific focus on this emerging 'next computing paradigm'. This shift recognizes the rapid evolution in the fields of technology and AI, and the strategic significance that these domains will take in global geopolitics. The document also highlights the intensified competition between leading global powers in technology and AI, emphasizing the implications for Europe. It underscores the need for Europe to seize the opportunities presented by these developments while mitigating risks, such as dependencies on external platforms and resources.

In the HiPEAC Vision 2024, the 'races' for leadership outlined in the HiPEAC Vision 2023 are viewed through the lens of NCP. These races include AI, innovative hardware solutions, cybersecurity, sustainability, and digital sovereignty. Given the cross-cutting importance of technological autonomy for Europe, the topic of sovereignty is now woven into the text, feeding directly into the recommendations of this HiPEAC Vision.

Structurally, the HiPEAC Vision 2024 retains the two-part format of previous editions:

1. A set of recommendations for the HiPEAC community at large, reflecting the shifts and new focus areas, with the NCP as the guiding topic.
2. A collection of expert articles divided into five key areas, each exploring a specific aspect of the race for technology leadership in the context of the NCP. Each of these five sections takes the position of Europe in that particular area into account. From this analysis, the recommendations are derived.

Readers are encouraged to engage both with this document, outlining the main recommendations, and with the accompanying HiPEAC Vision 2024 Rationale document, which contains the supporting articles, in order to fully understand the conclusions reached.

The full set of recommendations and articles may be found online:

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

Print editions of the HiPEAC Vision 2024, including the HiPEAC Vision 2024 Rationale document, are also available on request. Email [info@hipeac.net](mailto:info@hipeac.net) to request a copy.

<sup>1</sup> We are struggling for finding a good name to describe this 'next computing paradigm' (previously called the 'next web' in the HiPEAC Vision 2023). For a lack of better options, we will current stick with this name of 'next computing paradigm' or NCP. As a reminder of the acronym 'ICT', 'Intelligent Federated Computing and Technology' could also be an option. Please feel free to contact us at [info@hipeac.net](mailto:info@hipeac.net) if you have a good idea.

# HiPEAC Vision 2024

HIGH PERFORMANCE, EDGE AND CLOUD COMPUTING



## RATIONALE



**Editorial board:**

Marc Duranton (editor in chief), Paul Carpenter, Koen De Bosschere, Thomas Hoberg, Charles Robinson, Tullio Vardanega, Olivier Zendra.



Funded by  
the European Union

Supporting material for the HiPEAC Vision 2024 is available in the HiPEAC Vision 2024 Rationale document and online at [vision.hipeac.net](https://vision.hipeac.net)

# Contents

HiPEAC Vision 2024	6
Recommendations for the “next computing paradigm”	9
Recommendations for artificial intelligence	12
Recommendations for new hardware	14
Recommendations for cybersecurity	16
Recommendations for sustainability	18
Global recommendations	20
Acknowledgements	22
Process	23



# HiPEAC Vision 2024



For some years now, it has been obvious that technology is evolving faster than humans can adapt – witness, for example, the dizzying progress of artificial intelligence (AI) – while, at the same time, geopolitical and environmental pressures, such as global warming, are intensifying. Given this context, it is important that the European ecosystem of computing systems reacts quickly and efficiently to improve its place within the competition and proposes solutions that conform to “European” ethics.

This HiPEAC Vision sets out our vision of how Europe could do this: through the “**next computing paradigm**” (NCP), a convergence of key technologies from the web, cyber-physical systems (CPS), the cloud, the internet of things (IoT), digital twins, and AI into a coherent, federated ecosystem. This ecosystem emphasizes spatial computing, (generative) AI at the edge, dynamic web integration, user-centric models and intelligent, efficient and trustable orchestration of distributed services. Europe should work on building a coherent NCP and promote it as a world-wide approach.

Although based on existing technologies, the NCP will be highly disruptive. As the main highlight of the HiPEAC Vision 2024, it will rely on a synergy of the results of recommendations in the other topics outlined in the document.

As the HiPEAC Vision is now updated every year, some of the recommendations of the previous HiPEAC Vision necessarily still hold. Hence, for the 2024 edition, we

have retained the structure of the recommendations from the HiPEAC Vision 2023 as “leadership races”. These races, as we noted in the 2023 edition, are associated with the notion of urgency; the European Union (EU) needs to act fast. We are in a race against time because of external constraints, such as climate change, which is moving faster than we are [1], but also because technology is evolving very rapidly, and its potential impact is huge. One telling example is that of artificial intelligence. What large language models (LLMs) can do now was unthinkable just a few years ago. Indeed, the speed of technological evolution is too high to allow humans to adapt: education, ethics, laws, usage hardly can keep up with such a fast pace, although attempts have been made (such as the EU’s AI Act) [2].

There are also ongoing leadership races with other regions of the world, and not only in terms of economic interests. The fierce competition in LLMs between the big enterprises from the US and China, in which European companies are hardly represented<sup>1</sup>, has only intensified since the HiPEAC Vision 2023 was published. Similarly, as we noted previously, the race in the high-end semiconductor market and chip foundries has driven US, China, Japan and India to (re)build localized foundry capacity to ensure sovereignty. The European Union is also joining this race: the EU has developed its own Chips Act [3] to reduce dependence on other parts of the world in that strategic sector.

Another race which has only intensified since the last HiPEAC Vision is (cyber)

security, which is, more than ever, an arms race between attackers and professional security defenders. The number of cyber threats continues to grow, and large organizations are particularly vulnerable because they have more critical operations, more data to protect, and more money to pay a ransom.

Some technology races are advancing faster than others; for example, generative AI is now at the top of the hype curve (an annual update by Gartner), despite being practically unknown by the public at the beginning of 2022. The public release of ChatGPT in November 2022 prompted a drastic change with the fastest-ever adoption rate of a digital technology<sup>2</sup>. This generative AI technology will change many aspects of our daily lives, as the launch of the first iPhone did in 2007. Conversely, the “metaverse”, which was much hyped in 2022, is off the radar, although it will have its place in the future of computing.

Generative AI also impacts several topics that the previous HiPEAC Vision categorized into races: in **cybersecurity** attacks it can be the shield, but also the sword; jailbreaking LLMs to overcome their safety protection is more than a game. New **hardware** is needed to run those AI models economically, not only on large servers but also on embedded devices. Announcements have been made of smartphone integrated circuits (ICs) supporting foundation models<sup>3</sup>, the “core” of generative AI. Generative AI can also aid developments in the field of HiPEAC: there are already AI systems helping to produce, explain, and fix software code, and this same capability is also emerging for hardware.

<sup>1</sup> Or when a European company is present, it is often acquired by a non-European one.

<sup>2</sup> With more than 100 million weekly users.

<sup>3</sup> By Qualcomm [9] and Mediatek [10]

We could have chosen generative AI as the highlight of this edition of the HiPEAC Vision, but for the HiPEAC community it is only one element of a bigger picture: the “next computing paradigm” (NCP), mentioned above. We presented some aspects of this paradigm in the previous HiPEAC Vision, under the name of the “next web”. The NCP is the path we envision for future computing systems, both for software and for hardware. This theme is therefore the main highlight of the HiPEAC Vision 2024, and we will see how it will influence the other themes.

The six high-level recommendations are therefore:

1. Promote the “next computing paradigm” and develop the technologies that will make it happen.
2. Make the European Union a strong player in artificial intelligence, particularly for widespread use at the edge.
3. Develop innovative and efficient new hardware solutions, from architecture to technology.
4. Make cybersecurity a major upfront concern in every computing system.
5. Make sustainability lifecycle assessment a requirement for all new computing systems.
6. In all domains, foster global thinking, develop methods and tools relating to complex heterogeneous integrations and promote cross-domain / cross-topic collaborations.

A more detailed description of each high-level recommendation follows.

**Promote the next computing paradigm and develop the technologies that will make it happen**

The next computing paradigm entails an integration of the web, cyber-physical systems, cloud, IoT, digital twins, and AI into a coherent, federated ecosystem. It emphasizes spatial computing, (generative) AI at the edge, dynamic web integrations, user-centric models and intelligent, efficient and trustable orchestration of distributed services.

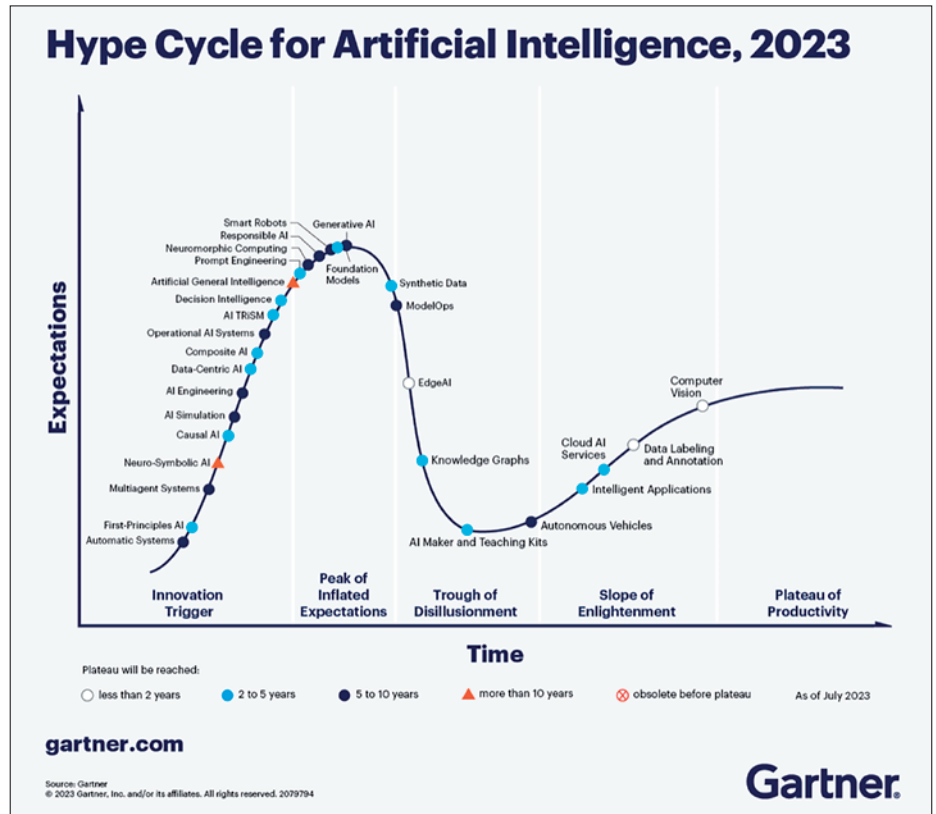


Figure 1: The Gartner hype curve 2023

**Make the EU a strong player in artificial intelligence, particularly for widespread use at the edge**

The focus on AI highlights the integration of LLMs in embedded devices (such as smartphones), the capabilities of various processors to run these models, the role of smaller LLMs in smart devices, and their societal and technological impact, including for the HiPEAC community.

**Develop innovative and efficient new hardware solutions, from architecture to technology**

Recommendations for new hardware development include proactive architecture development, fast chip / system prototyping, a global value-chain perspective, developing the chiplet / interposer ecosystem, and a focus on sustainability and sovereignty in hardware development.

**Make cybersecurity a major upfront concern in every computing system**

The cybersecurity recommendations deal with the challenges associated with the NCP and offer directions for solutions, both

in hardware and software, including the development of security tools, of vulnerability identification and mitigation, of open-source and EU-made resources that can be audited, the automation of cyber-defence tools, and the enhancement of user authentication and privacy control. They also highlight the impact of LLMs, which can be used to detect and protect against malicious attacks, but can also be used to engineer such attacks. LLMs are also liable to attacks that bypass the security protections embedded in their structure to avoid responding to unethical questions.

**Make sustainability lifecycle assessment a requirement for all new computing systems**

Sustainability recommendations include designing lifecycle models for information and communication technology (ICT) [4] devices and services, developing a sustainability-focused design methodology, new business models considering the full lifecycle, and developing effective “ICT for green applications”.



### **In all domains, foster global thinking and promote cross-domain / cross-topic collaborations**

The global **recommendations focus on promoting collaboration across various technological domains, launching cross-domain** project calls, and establishing competence centres. Emphasis is also placed on leveraging open-source hardware and software, and developing tools to address multi-dimensional challenges. We should continue to bridge and “break silos” so that we improve EU capacity to create and develop innovations in the computing (continuum) domain.

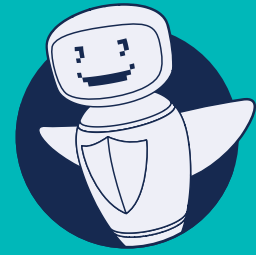
It has become increasingly obvious that digital technology in general, and the web in particular, constitute critical infrastructure, and that control over that infrastructure has major implications for geopolitics. Given the cross-cutting importance of technological autonomy for the EU, the topic of sovereignty is now more focused and is included in each key area.

This section of the HiPEAC Vision 2024 includes a discussion of the main recommendations with short explanations. We encourage readers to look at the more detailed “Rationale” document, which is structured according to the main themes of the HiPEAC Vision 2024 and which

includes articles that explain the rationale behind the HiPEAC Vision recommendations in more detail, helping readers understand the current situation, the forecast of things to come, and the reasons why this set of recommendations was formulated. The sections and their associated articles can be read independently and are available at [vision.hipeac.net](http://vision.hipeac.net) [5].

Each high-level recommendation is discussed in the sections that follow in an extended overview, which includes the basis for each recommendation and the motivation behind its inclusion in the HiPEAC Vision 2024.

# Recommendations for the next computing paradigm



Promote the “next computing paradigm” and develop the technologies that will make it happen



Figure 2: Picture generated by Dall-E 3

The digital environment based on computing systems is undergoing profound change, and we are beginning to see what it might look like in the future, merging the technologies of the web, cyber-physical systems (CPS), the cloud, the internet of

things (IoT), digital twins, the computing continuum, the metaverse and AI. This is based on the following observations (illustrated in Figure 4):

- 1) Increasing quantities of data are being generated by interconnected computing machines to create new services. There will be an integration of the “web of machines” with the “web of humans”.
- 2) Computing systems will be increasingly linked to the physical world, as interconnected machines act in the real world, either by analysing it to provide data, or by acting on it. Real-world constraints such as security, real time and localization will have to be taken into account. The future digital infrastructure (augmented by digital twins – which enable time to be considered as a complete variable, i.e. to model the past, the present and the future) will be the backbone of industry and most of the digital world.
- 3) In addition to time, spatial aspects (location, the physical composition of the world) will increasingly be taken into account. For example, autonomous vehicles or drones must take into

account the 3D aspect of the world in order to navigate, and the rules do not apply identically regardless of location (a drone cannot fly over certain areas, a car cannot enter stores, etc.). On the other hand, “realistic” visualization of calculated results (by virtual, augmented or mixed reality) will play an increasingly important role (the metaverse is one expression of this). Computing systems will therefore increasingly need to use and provide spatial (and temporal, hence at least 4D) services.

- 4) The key mechanism: “everything as a service” (XaaS). Applications will be made up of a composition of local or remote services that are dynamically aggregated to provide functionalities (what we currently call a program – we are starting to see this in Python programming: Python is growing into being just an integrator, an orchestrator of functions (=services) provided by external specialized libraries). These services (including digital twins) will operate where it is most efficient, according to the user’s criteria, which could be security (locality), response time, cost, accuracy, etc. These services will therefore have to be infinitely composable, and offer information not only on their functionality, but also on their non-functional properties, such as latency, ecological impact, consumption, cost, and so on. For example, a temperature sensor offers the service of providing a temperature at a certain place, on a certain date (hence 4D information) with certain accuracy and energy consumption, etc.

- 5) Intelligent orchestrators will be needed to find, select and compose services to perform the required task. They will also have a role to play in protecting

## Key points 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

- The “NCP” will be increasingly intertwined with the physical world. Real-world constraints such as safety, real time and localization will need to be considered. The “NCP” (augmented by digital twins - to model past, present and future) will be the backbone of industry and most of the digital world.

### The spatial web

- The “NCP” will also be spatial: we will go/are into the computing continuum (ubiquitous computing). The metaverse is one expression of this.

### Everything as a service

- Key mechanism: “everything as a service” (XaaS). Applications for users 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 for this to protect and guide users into the complexity of services. This will use AI techniques for smooth and natural communication with humans (e.g. LLMs).



Figure 3: Key points for the NCP

and guiding users through the complexity of services. For this, they will need to be trusted. They will use AI techniques for smooth and natural communication with humans (e.g. LLMs).

This envisaged evolution of computing systems is based on elements offered by current systems, but also offers numerous research challenges: for example, how to ensure integrations and interoperability; how to present the characteristics of each service so that it can be soundly discovered and chosen; how to ensure that what it presents can be trusted; how to ensure that these services are executed where they are most efficient (this entails moving not only data, but also code) and on any architecture (the instruction set – x86, ARM, RISC-V – becoming irrelevant with approaches such as WebAssembly, for example); how to protect against attacks (cybersecurity); how to ensure expected levels of performance; how to take into account other non-functional properties such as time, latency, localization, privacy; and so on.

The following recommendations can be set to trigger the transition toward this next computing paradigm (NCP):

**1. Develop 4D-aware implementation technologies**

The concept of 4D computing (in the three dimensions of space plus time) is not

merely theoretical; it has practical implications across various sectors. Consider smart homes that adjust to the occupants’ routines or supply chains that utilize predictive analytics for inventory management based on real-time and historical data. These use cases require a shift from static computational models to dynamic ones that are context aware and time sensitive. Time is also important to fulfil the requirement of interacting or driving physical devices in a timely fashion: current web infrastructure primarily operates on HTTP, which is not optimized for the low-latency requirements of 4D data. Upgrading web protocols to accommodate the spatial web will involve incorporating standards like WebRTC for real-time communication, allowing instantaneous interactions for applications such as telesurgery, where even milliseconds matter.

Hence, the resulting recommendation is to develop a standard representation and suite of protocols for encoding properties of physical objects and spaces, supporting mobile computation and devices, and enabling 4D-enabled operations. For example, the focus on spatial and time-aware computing could be achieved by merging evolutionary versions of HTTP-based REST-like web protocols with spatial web protocols (HSML and HSTP, OpenUSD). This involves ensuring timely delivery for any granularity of time.

**2. Augment the APIs of services for interoperability**

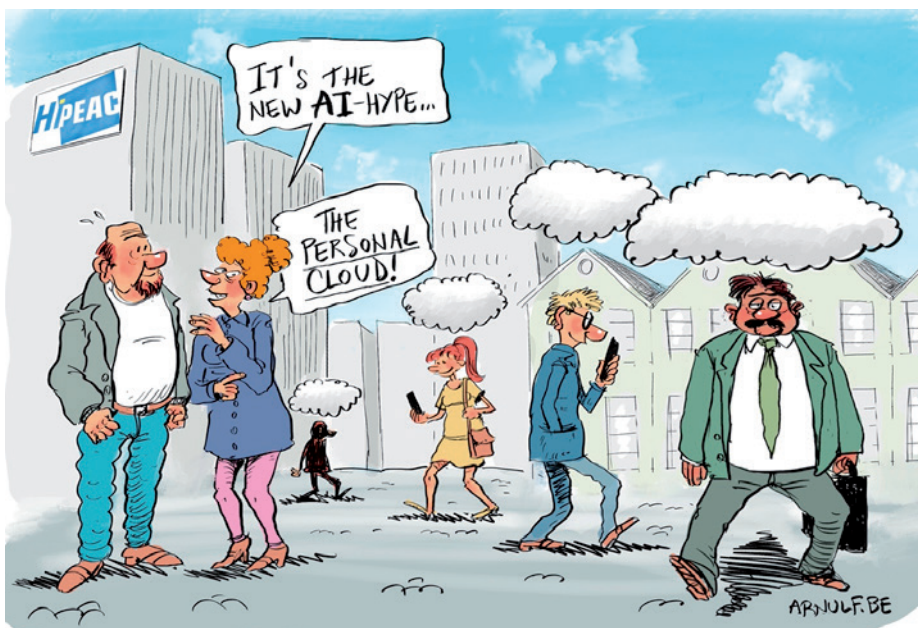
The APIs of interconnected services will need to be enhanced with contract-based specifications, akin to service-level agreements (SLAs), ensuring services can communicate with an expected level of performance. This assume / guarantee model means that an API not only promises to deliver a service but also specifies the conditions under which it can perform optimally.

The associated recommendation is to enhance APIs with contract-based interoperable specifications to enable ephemeral compositions based on the pairing of assume / guarantee declarations across interfaces. This will allow for dynamic and efficient compositions of services and data.

**3. Enable the mobility of computation**

The mobility of computation involves dynamically relocating data and processes closer to where they are needed. This could take the form of “follow-me” data (and execution) environments where a user’s preferences and settings are instantly available on any device they use, by computing “on edge devices” — closer to the user. Spatial computing will expand to support dynamic aggregations of compute nodes. On demand (of a smart orchestrator), selected services can create an on-the-fly mesh network of computing resources to manage logistics, resources, and communications effectively, adapting to the environment to deliver the best composition of services. This allows the integration of all elements into a seamless distributed and networked platform where resources are accessed using streamlined web-level protocols, and where application services can be dynamically aggregated.

WebAssembly (Wasm) offers a way to run code written in multiple languages on the web at near-native speed. By further developing Wasm, computationally intensive tasks like image processing can be executed faster and more securely within web applications, vastly improving user experience and opening new possibilities for web-based applications, enabling computation to move after travelling users



or objects, or towards specific data sources. This aims to balance factors like latency, privacy, data freshness, and energy requirements. Improvements in WebAssembly-type technology are essential for hosting and for the efficient execution of in-transit computations.

**4. Adopt (generative) AI at the edge**

Embedding AI engines in edge devices allows applications that operate more efficiently and with reduced latency and less communication overhead, while ensuring greater privacy. AI at the edge will also be a natural interface for users. Therefore, our recommendation is to embed personalized AI engines in edge devices and to develop robust, decentralized AI models.

**5. Develop AI-powered edge-based orchestrators**

AI-based orchestrators at the edge can dynamically combine different services for dynamic, on-the-fly construction of smart, ad-hoc orchestrated applications in response to users' requirements. In smart transportation systems, this could mean real-time re-routing of traffic based on current conditions and predictive analysis, optimizing flow, and reducing congestion, all without human intervention. Smart orchestrators are enablers for ephemeral, context-aware aggregations based on web resources to improve accessibility

and efficiency. These orchestrators, robust and trustworthy, should dynamically and opportunistically assemble remote APIs into ad-hoc, private services. To cope with privacy and efficiency, they should be at the edge, near the user: it is important in the future to move towards computing models that place users or data sources at their core. This ensures the balance of latency, privacy, and energy requirements, and has implications for system design and user experience.

**6. Encourage non-proprietary integration**

The integration of non-proprietary technologies is crucial for fostering innovation. By supporting open standards and platforms, developers can create cross-compatible tools and applications that benefit from community improvements. In academia, this would allow for a collaborative approach to research, leveraging shared tools and datasets to accelerate discovery. The EU should be at the forefront of development of new open standards and technologies.

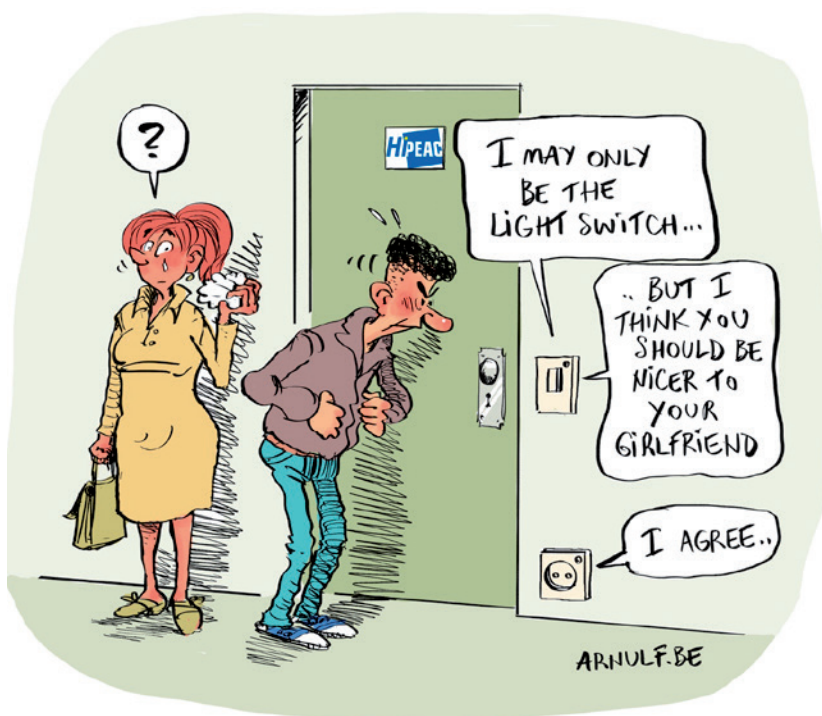
**7. Initiate proof-of-concept level efforts, demonstrative testbeds and business cases**

The NCP will only appear if users see its potential and advantages, and if there are clear business cases and drive. It is

therefore necessary to develop proof-of-concept demonstrators so that users and business can assess the advantages of this combination of multiple technologies. It also helps researchers and integrators to discover problems on real-life uses. The first feasibility demonstration could be testbeds of "trusted orchestrators", running on any mobile device and appliance to serve the needs of safe, private, personalized and effective service composition on the computing continuum.

**Closing remarks**

The recommendations laid out for the NCP have the potential to create a more connected, responsive, and intelligent world. As we move towards this new horizon, the focus should not only be on technological capabilities but also on creating an inclusive and sustainable ecosystem that benefits all stakeholders. The vision for the next generation of computing is one of unlimited potential, where the barriers of time and space are minimized, and the digital and physical realms converge in ways that are currently unimaginable. HiPEAC envisions a transformative future of computing that integrates digital and physical realms, spatial awareness, and generative AI. It calls for collaborative efforts across multiple domains to realize this vision of a more efficient, context-aware, and user-centric digital ecosystem.



# Recommendations for artificial intelligence



## Make the EU a strong player in artificial intelligence, particularly for widespread use at the edge

Artificial intelligence (AI) represents a frontier for innovation and has become a pivotal point of strategic competition globally. As Europe navigates the challenges and opportunities presented by AI, particularly large language models (LLMs), it is imperative to adopt a coherent and proactive strategy that ensures growth, leadership, and the ethical deployment of these technologies.

HiPEAC outlines key recommendations for EU's approach to AI, with a focus on fostering a robust AI ecosystem that is inclusive, competitive, and aligned with "European" values. The recommendations proposed in the HiPEAC Vision 2023 are still valid and now more urgent due to the explosion in use and potential risks trig-

gered by giving the public access to ChatGPT in November 2022, an event that has significantly boosted the impact of AI technologies.

### 1. Invest in AI core technologies and development across application domains to support growth for Europe

The rapid expansion of AI, especially LLMs, into diverse application domains will be enhanced by a focus on multimodality. By integrating text, image, and sound, AI can achieve more comprehensive understanding and generate more nuanced responses. The EU must invest in research and infrastructure to support multimodal AI development, ensuring its applicability across sectors such as healthcare, education, and public services, thereby fostering an environment of innovation and practical AI utility.

### 2. Develop and provide access to foundational models that support "European" values

For the EU to gain a competitive edge on the global stage, it is essential not only to participate in but also to lead the development of foundational AI models<sup>5</sup>. This involves creating and sharing methodologies and datasets for fine-tuning these models to suit specific regional needs. By doing so, Europe can secure its sovereignty in AI technologies and promote a digital economy that reflects its standards and ethics. The foundation model Bloom [8], created via the open-science group "BigScience", is an example of initiatives that should be further promoted. Bloom is able to output coherent text in 46 languages and 13 programming languages and was trained between March and July 2022 (before the public announcement of ChatGPT). It was trained on the public supercomputer Jean Zay, provided by the French government and is recognized as a foundation model with one of the lowest CO<sub>2</sub> footprints for its size (176 B parameters).

### 3. Promote open-source models

Open-source models serve as the backbone of a collaborative AI ecosystem. Platforms like HuggingFace democratize AI by providing access to state-of-the-art models and fostering an environment where researchers and practitioners can contribute and innovate. Europe should encourage the growth of such platforms, ensuring open access to AI resources and facilitating a culture of shared progress.



<sup>5</sup> The company Mistral.ai is a good example; its models are available for download (although not the training data set) and are very well ranked for their performance.

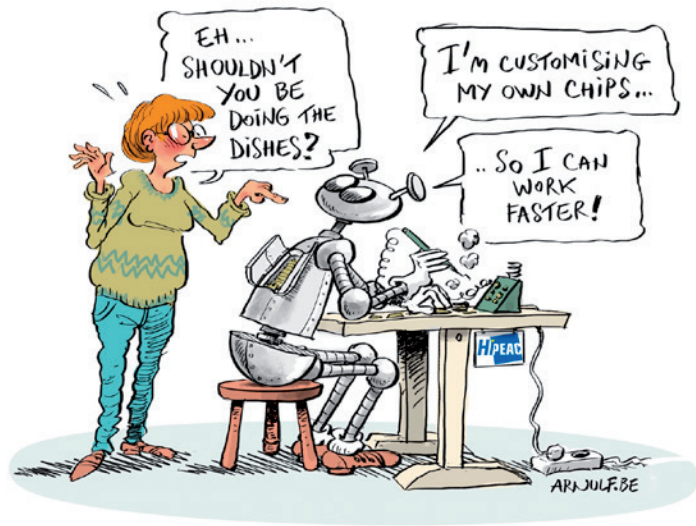
**4. Develop local solutions and specialized accelerators for the integration of LLMs in smart devices**

LLMs will be in smartphones by 2024, driven by advancements in processors like Qualcomm Snapdragon 8 Gen 3 [9] and MediaTek Dimensity 9300 [10]. This integration represents a significant leap in the computational capabilities of mobile devices, enabling them to run sophisticated AI models locally. Together with the increasing power and efficiency of smaller LLM models, this shows a trend of making AI based on LLMs suitable for integration into smart devices at the edge. These advances suggest a future where smart devices can perform complex AI tasks without relying on cloud-based computing.

Developing localized solutions and specialized accelerators for running LLMs at the edge is pivotal. This empowers real-time AI applications on devices, reducing dependency on centralized data centres, and enhancing privacy and efficiency. Europe’s strategy should include support both for start-ups and for established companies in developing edge AI capabilities, fostering a decentralized and resilient AI infrastructure. AI running on edge devices is also a key ingredient to realize the NCP.

**5. Use AI to aid software and hardware development**

With a shortage of specialists in high-performance computing, edge and cloud – i.e. HiPEAC – domains, LLMs can be a significant aid to software development, while also helping to improve productivity in developing efficient hardware. By upskilling current engineers and researchers, Europe can amplify their productivity with AI, leveraging these technologies to address complex computational problems and accelerate the development cycle. The concept of “centaurs”, or partnerships between AI and developers, is a promising approach to enhance productivity and code quality. Europe should invest in tools and platforms that facilitate this symbiosis, enabling developers to harness AI for more efficient and creative development processes.



**6. Continue to develop policies around accessibility and societal impact**

The societal impact of AI cannot be overstated, and its accessibility is a cornerstone of digital inclusivity. AI technologies should be available to all European citizens, ensuring that the benefits of digital transformation are equitably distributed. This requires policy frameworks that promote accessibility, education, and ethical AI practices, working towards a society where technology serves the public good.

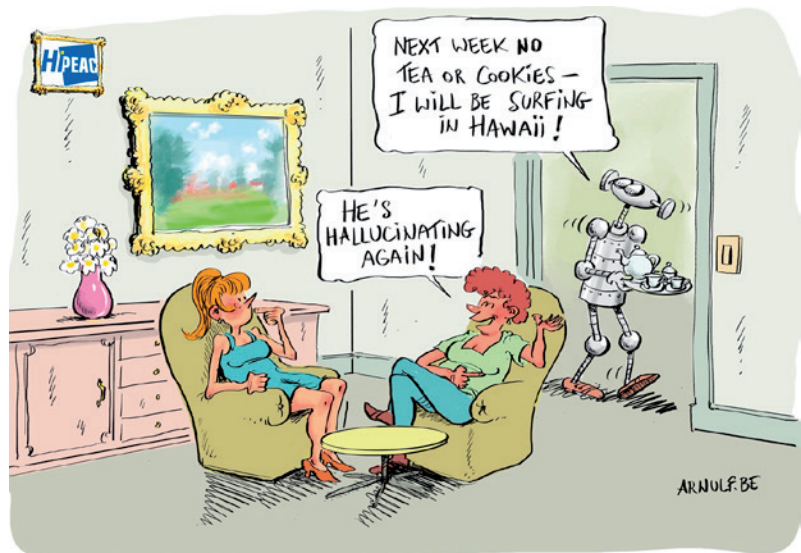
**7. Ensure “correctness by construction” in AI models**

Ensuring that the code and hardware generated by AI are “correct by construction” is a critical aspect of trust and reliability. Making sure that AI models generate sound and validated answers (and do not “hallucinate”) is key to ensure that AI can be effectively used to carry out the previous recommendation on accessibility and social impact. Europe should lead the

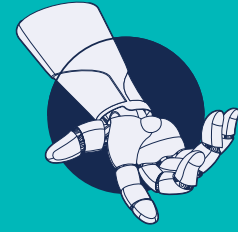
way in developing methods that automatically verify the correctness of AI-generated outputs, thus reducing the need for extensive human oversight and increasing the trustworthiness of AI systems.

**Closing remarks**

The strategic approach to AI outlined in these recommendations presents a roadmap for Europe to navigate the AI landscape with confidence and vision. By embracing these initiatives – for example, developing “made in Europe” foundational models and promoting open science – Europe can foster an AI ecosystem that is not only competitive but also reflective of the EU’s commitment to open innovation, ethical standards, and societal wellbeing. To achieve this, and thereby to realize the full potential of AI as a force for good in European society, it will be necessary to coordinate action across all levels of government, industry, and academia.



# Recommendations for new hardware



## Develop innovative and efficient new hardware solutions, from architecture to technology

There have been no major changes in this field since the HiPEAC Vision 2023 was published, so the previous recommendations still hold. The boom of AI is increasing the need for efficient accelerators, while, conversely, as explained in recommendation 5 of the AI section, AI can also be used to facilitate the exploration of the design space and help hardware developers. The modular approach based on chiplets and interposers has bridged the gap from research to market, and is now in use by major chip manufacturers. However, the ecosystem is not fully developed yet, and standardization will be key to ensure that system integrators (when building systems using interposers) will be able to use chiplets from various providers. We are not at that point yet, and Europe should promote the creation of this ecosystem to allow its companies and start-up to differentiate and to offer efficient solutions for different use cases.

Therefore, the proposed recommendations are as follows:

### 1. Continue the quest for greater performance and lower energy use

HiPEAC emphasizes the ongoing quest for hardware that offers greater performance and lower energy usage. This is particularly important as Moore's Law slows, necessitating innovative approaches to maintain technological progress. Exploration of promising technologies (photonics, using physical phenomena to do computations, storage) and innovative architectures are research domains that should continue to prepare the future.

HiPEAC encourages emerging technologies in hardware, including advances in memory technologies like MRAM and spintronics, progress in silicon technology such as stacked nanosheet FETs and photonics for computing, and the evolution of organic and flexible electronics (for particular use cases).

### 2. Promote interdisciplinary research

Interdisciplinary research is highlighted as a key driver for new hardware paradigms. This includes, for example, exploring quantum computing, neuromorphic computing, new materials and information theory, each offering unique contributions to hardware development. This should be complemented by close discussion with the system architects to define where each topic is the most effective and the software community in order to deliver solutions that are efficiently usable for end users.

### 3. Explore innovative architectures

Innovative architectures are being driven by the demands of data-intensive computing. These include in-memory computing, 3D stacking, and using adapted precision for accuracy, each tailored to specific application needs. Europe should be proactively involved in defining new architectures and accelerators.

### 4. Develop a complete ecosystem to help innovative European solutions mature and gain traction

HiPEAC promotes solutions that encourage fast chip and system prototyping, especially within academia and start-ups. It is important to adopt a global perspective on the complete value chain, focusing on specialized architectures, system intellectual properties (IPs), electronic design automation (EDA) tools, and the development of

chiplet-plus-interposer ecosystems. A large set of IPs are required to make a system on chip (SoC): not only an open instruction set like the one proposed by RISC-V, but also architectures that implement it, coprocessors, memory interfaces, peripherals, etc. These additional IPs are complex and are rarely offered as open-source hardware. Furthermore, all the software tools to help define the architecture, the IP libraries and low-level libraries that make the link with the technology of the foundry, the EDA tools to assemble, verify and test a SoC before tape-out are also often proprietary, and tend to be very expensive. Access to foundries should be also facilitated, together with the test and packaging of the chips.

It is also important to develop an industrial ecosystem of interoperable chiplets, so that integrators (building systems, and defining interposers), are able to use a large variety of chiplets from various providers. This will create a brand-new ecosystem, which will promote innovation and in which Europe can be a leader.

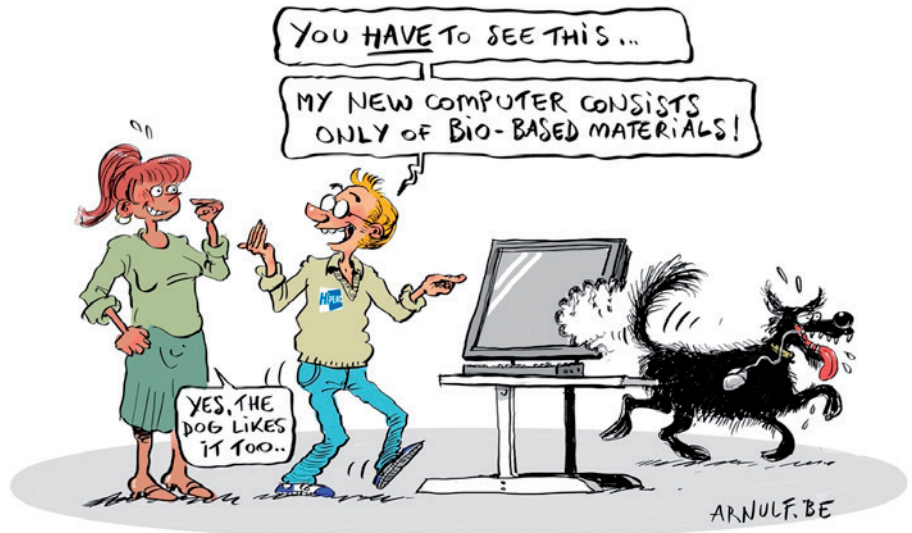
On 30 November 2023, the European Commission officially launched the Chips Joint Undertaking under the European Chips Act [3], which came into force on 21 September 2023. With a budget of €11 billion between 2023 and 2030 from the European Union and member states, this initiative aims to bridge the gap between research and industry by facilitating the industrialization and commercialization of chip innovations. This initiative is a step towards increased innovation in Europe and covers some of the above-mentioned topics.

## 5. Reduce the ecological impact of electronic devices

The ecological impact of making electronic devices should be reduced. While most of it is not covered in the scope of HiPEAC (i.e. silicon process and fabrication), the HiPEAC community should look for innovative solutions that promote sustainability in hardware and software development, addressing energy consumption, CO<sub>2</sub> emissions, and material usage, and exploring innovative concepts like flexible / edible electronics.

### Closing remarks

With its emphasis on AI at the edge and interconnection, the NCP will continue the drive for more efficient hardware solutions that can process massive amounts of data on a low energy budget, drawing on concepts such as data-centric architectures. The EU has taken steps to take control over parts of the hardware value chain, and these should be continued to ensure Europe's technology position on the global stage.





# Recommendations for cybersecurity



## Make cybersecurity a major upfront concern in every computing system

Cybersecurity, in all its aspects, is an increasingly important preoccupation in the ICT domain. The battle between the sword (attacks) and the shield (protection) is intensifying, up to the level of actual cyberwar between countries. As computing systems are increasingly used to control more and more devices, and these devices are often interconnected, the consequences of attacks are all the more devastating. Closing hospitals and public services, stopping factories etc., are just the beginning. Attacks on autonomous vehicles or on critical infrastructure can put lives at risk. The proposed next computing paradigm, relying on collaboration between smart orchestrators and services, should be built from the outset to prevent rogue or fake services and malicious use as far as possible. Security, authentication and verification should be present at every level to ensure that the next computing paradigm delivers its promises, even in the face of attacks from hackers or malevolent participants.

The global dissemination of LLMs also impacts cybersecurity: in addition to deep-fakes, making it easier than ever to create a false reality or identities, they are also tools that, when used maliciously, can lead to new cyberattacks, at all levels. However, LLMs can also be used as shields, analysing and detecting anomalies to protect against attacks. The safety checks built in LLMs are far from exhaustive; there is even competition in the field of on how to “jailbreak” them, and hence unleash their use for malevolent purposes.

Our key recommendations include prioritizing security and privacy in NCP

design and implementation, both for hardware and software components, developing tools for vulnerability detection and mitigation, emphasizing relying on open-source and EU-made resources for critical NCP components, implementing automated cyber-defence tools, strengthening authentication processes, and establishing robust certification and liability measures for ICT system providers.

Therefore, in addition to the HiPEAC Vision 2023 cybersecurity recommendations, which still hold, HiPEAC proposes the following aggregated recommendations in the scope of the NCP:

### 1. Recognize and protect vulnerabilities in a connected society

HiPEAC highlights the broad attack surface created by the ubiquity of ICT systems in our civilization (which will be also the case with the NCP). Privacy is constantly threatened, or not respected

(e.g. in browsers). Cyberattacks present a severe threat, potentially causing economic devastation and harm to individuals. The growing sophistication and automation of cybercrime increases the magnitude of the issues. Microarchitectural security flaws are also a major concern, being particularly difficult to spot and counter.

All systems, both hardware and software, should be built with cybersecurity concerns ingrained as early as possible in their design and implementation. Separation of concerns and isolation of elements could be a solution to contain the dissemination of malware. This is one case where silos are, in fact, useful. The integration of multiple components and specific hardware-software interactions in the NCP create further vulnerabilities. The rapid adoption of new technologies like LLMs, often without thorough security assessment, increases the risks.



## 2. Reduce dependency on external ICT resources

Many hardware and software components integral to the NCP could be proprietary and often sourced from outside the EU, leading to dependencies that have technical, economic, political and even military consequences. These dependencies pose a risk to the security and sovereignty of ICT systems, in addition to making auditing difficult. Trusted European companies or open-source solutions should provide both the design and the production of the most essential elements that will build the NCP.

## 3. Address security challenges in LLMs

The availability to the public of LLMs, exemplified by ChatGPT, also triggered the existing battle between sword and shield, but at several levels:

- LLMs can be used to help evaluate and prevent cybersecurity breaches by analysing code and systems.

- However, they can also be used by hackers to develop new, easy-to-use ways to exploit vulnerabilities (or find out how to make bombs, etc). As most LLMs have protections against malevolent use, this leads to hackers trying to find ways to bypass the LLMs' protections, "jail-breaking" them by using more and more sophisticated prompt injection, leading to the output of "uncensored" data.

As the nature of LLMs means that the information is not explicit in an identifiable part of the model, it is very difficult to "erase" sensitive information. Of course, they can be trained with curated datasets without "dangerous" data, but the size of the dataset, and the curation, which is mainly done by humans, make this virtually impossible or too expensive. Fine tuning and other more explicit approaches to sort the input and output are being developed, but they cannot 100% guaran-

tee that the LLMs are "safe" and that they will not generate output that can be used by malevolent people. Therefore, if the LLM technology is widely available, it is important to develop new techniques to increase their safety, at least for the large foundation models interacting with a large audience.

## Closing remarks

The NCP proposed by HiPEAC embodies an even more connected and distributed world, where cyber-physical systems play a more important role, and where AI is omnipresent. Due to its distributed architecture, it is no longer the case that there is one central point of attack, but robust authentication, checking and redundancies of data and services should allow a higher level of protection. Ensuring that cybersecurity is addressed from the outset is therefore pivotal, given the stakes involved if systems are attacked.

# Recommendations for sustainability



## Make sustainability lifecycle assessment a requirement for all new computing systems

To address the profound impact of humanity on Planet Earth, it is imperative that computing systems either minimize their environmental footprint (“Green ICT”) or actively contribute to reducing the ecological impact of other industries (“ICT for green”).

On one hand, the computing industry has made significant strides in reducing energy consumption over several decades. For instance, in the past three decades, the energy required per floating-point operation has decreased by a staggering factor of approximately 100,000,000 – a feat unmatched by any other sector. On the other hand, the escalating demand for computational power has outpaced

these energy efficiency gains, resulting in a continual rise in global energy consumption within the ICT infrastructure.

Huge though this energy consumption may be, it is essential to recognize this represents just one facet of the ecological footprint associated with computing. Equally significant is the embodied energy of computing devices, which often surpasses the operational energy consumed throughout their entire lifespan<sup>6</sup>.

In order to contribute to global sustainability goals, it is crucial to consider the complete lifecycle of a device, encompassing production, the bill of materials required to construct a device, shipping, usage, end-of-life processing, and more. The overarching objective should be the minimization of the total impact.

At present, the primary challenge lies in the absence of universally agreed-upon models for conducting lifecycle assessments of the devices and services offered by the ICT industry. The inability to reliably calculate the ecological footprint hampers the integration of this information into design decisions, leaving the ICT industry without a clear direction. Therefore, the most pressing recommendation in terms of sustainability is to focus on developing lifecycle models.

The HiPEAC sustainability recommendations are as follows:

### 1. Design and improve validated lifecycle models

HiPEAC strongly recommends the urgent development of precise and universally accepted lifecycle models for ICT devices and services. These models should not only encompass factors like embodied and operational energy but should also comprehensively account for the ecological impact of activities such as mining, water usage, the use of chemicals in production, and the environmental consequences at the end of a product’s life cycle. HiPEAC believes that the EU Digital Product Passport [11] is a good tool to facilitate access to and exchange of this critical information.

Furthermore, HiPEAC proposes the creation of sustainability labels designed to assist consumers in making informed choices and increasing awareness about sustainability. Achieving this level of transparency and standardization will require the establishment of a universally accepted method for determining the environmental impact of devices and services.



<sup>6</sup> For instance, in the case of modern smartphones, approximately 85% of their environmental impact is attributed to embodied energy, while only 15% relates to operational energy usage.

### 2. Develop sustainability-focused design methodologies

HiPEAC emphasizes the importance of integrating the lifecycle ecological footprint as a fundamental design objective for future systems. By making the ecological footprint a central consideration in the design process, designers will naturally take into account factors like repairability, reusability potential, recyclability, and the effective management of end-of-life processing right from the initial stages of product development. It is essential to highlight that the successful implementation of these design practices relies heavily on the availability of universally accepted and reliable lifecycle models.

### 3. Create new business models incorporating the full lifecycle

Meeting sustainability goals requires companies to embrace a different approach,

one that goes beyond production and encompasses the maintenance of products long after their initial sale. This shift requires a comprehensive re-evaluation of every facet of a company's operations, including its mission, product development, marketing, sales strategies, customer relationships, and even human-resources practices. HiPEAC underscores that the absence of viable sustainable business models is a significant barrier to meaningful actions by companies aimed at enhancing the sustainability of their products. Therefore, rethinking business models is crucial to drive positive change in the sustainability landscape.

### 4. Prioritize ICT for green applications

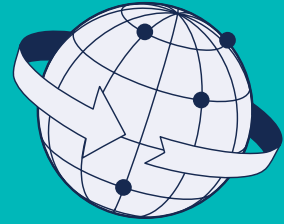
In addition to developing more sustainable digital technologies, HiPEAC recommends that the community focuses on developing "ICT for green applications",

and works collectively with various domains in investigating how ICT technology could reduce the environmental impact of key processes in these domains. Examples include smart grid optimization, precision agriculture, and electromobility.

### Concluding remarks

There is a very long way to go to delivering sustainable electronics. We believe that the NCP offers significant advantages for users, including simplified, more natural interaction with computing systems, enhanced privacy and more robust safety protections. However, this vision will only be realized if the NCP's components and systems are delivered in an environmentally responsible way.

# Global recommendations



## In all domains, think globally and promote cross-domain / cross-topic collaborations

The global recommendations of HiPEAC Vision 2023 still hold in 2024, so the reader can also refer to this part of the HiPEAC Vision 2023. This “global thinking” is also a key ingredient for building the NCP.

However, since the publication of the HiPEAC Vision 2023, progress has been made in some areas:

- HiPEAC organized a meeting between the major organizations / initiatives in its field (ETP4HPC, AIOTI, BDVA, ADRA, Chips JU, 6GSNS, NESSI, ECSO, the FIWARE Foundation, and Destination Earth), where it was decided to create a “computing continuum” action [6] as

an exchange forum between these associations in order to synchronize their findings and recommendations on the “next” steps and on their SRAs (Strategic Research Agendas), SRIAs (Strategic Research and Innovation Agendas), roadmaps and other forecasting documents.

- Progress in the field of AI, especially on LLMs, has increased the practical feasibility of smart orchestrators (LLMs can automatically generate “glue” code to orchestrate application programming interface (API) calls, as demonstrated by ChatGPT’s ability to use Python libraries to execute particular tasks).
- A number of EU-funded research and innovation projects are tackling issues related to the priorities outlined below, including orchestration and / or meta operating systems for the edge-cloud-IoT continuum.

Hence, the summary of the global recommendations is:

### 1. Promote collaboration

HiPEAC recommends promoting collaboration across research-and-technology teams and organizations in Europe. Initiatives like the TransContinuum Initiative [7] and the Computing Continuum<sup>4</sup> exemplify such collaborative frameworks, aiming to create synergies across various technological domains.

### 2. Launch cross-domain project calls

HiPEAC recommends the launch of cross-domain project calls to foster synergies between application and technology domains. This approach is designed to encourage interdisciplinary research and development, leading to innovative solutions.

### 3. Establish competence centres

HiPEAC recommends establishing cross-disciplinary EU competence centres to centralize expertise and resources (à la CERN). These centres would facilitate knowledge exchange and streamline research efforts across diverse technological fields. They could have a physical location or be virtual, but in either case should inspire active participation by experts.

### 4. Leverage open source

HiPEAC recommends promoting and leveraging open-source hardware and software. Open-source ecosystems can accelerate innovation, foster collaboration, and ensure wider accessibility to technological advancements. The code can also be audited for security or other requirements.



<sup>4</sup> The new action initiated by HiPEAC to promote exchange between organizations like ETP4HPC, AIOTI, BDVA, ADRA, the Chips JU, 6GSNS, NESSI, ECSO, the FIWARE Foundation, etc.

**5. Develop tooling for multi-dimensional challenges**

HiPEAC recommends developing support capacities that pull together stakeholders to address cross-disciplinary and multi-dimensional challenges. Also, utilizing digital twins and AI as tools can aid in cross-engineering activities by proposing innovative solutions and

optimizing system performance. They can also form a bridge between different parts of an organization (for example, a digital twin of an electronic system can be used by a mechanical engineer to check how it could fit into the system, and software developers can test their software before the hardware is ready, etc.).

**6. Take a holistic approach to efficiency**

HiPEAC stresses the need for a holistic approach and global co-design system thinking. This involves optimizing global efficiency by finding synergies between various domains, bridging and breaking traditional silos in technology and application areas.

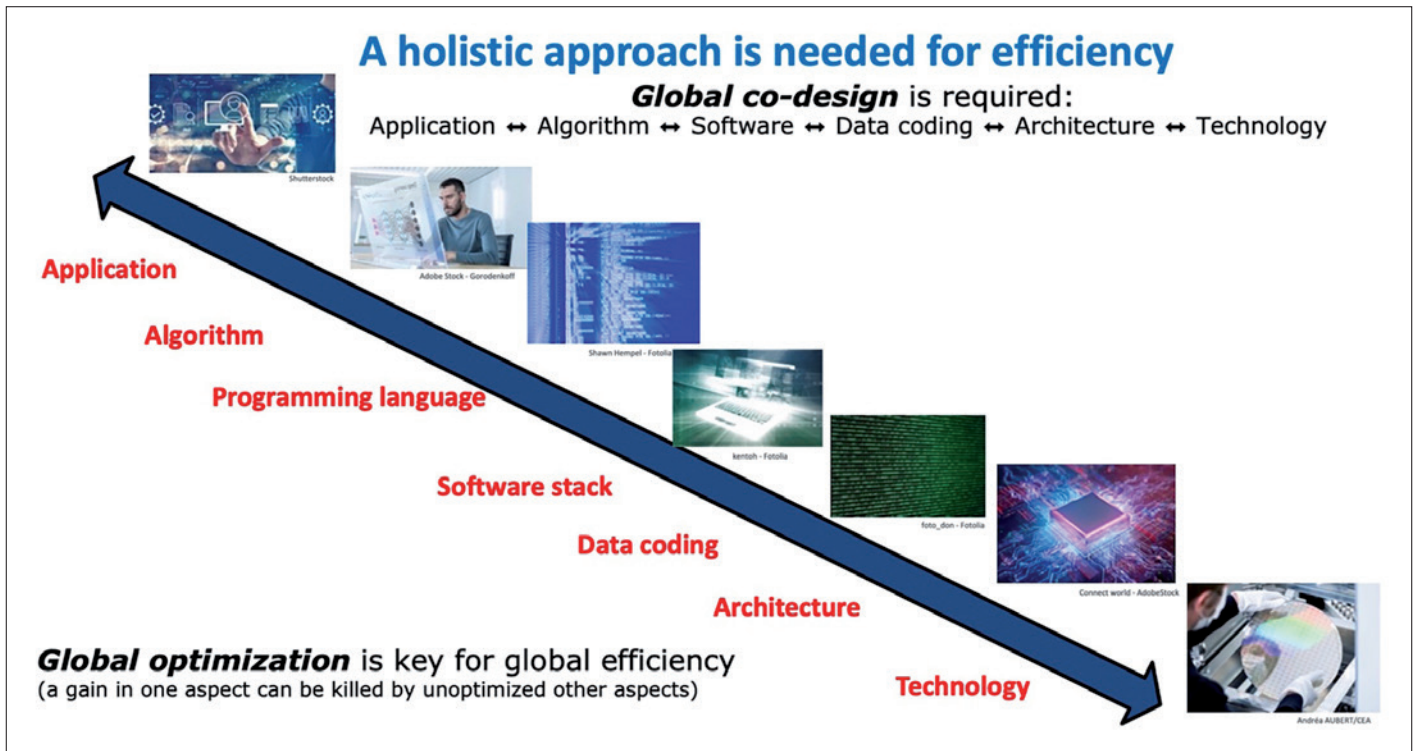


Figure 4: the sum of local optimizations is less efficient than global optimization. Each of applications, algorithms that are executed in the application, the programming language in which the algorithms are implemented, the software stack (compiler, operating system, middleware, libraries) that result from the mapping of the algorithm into a language that is compiled or executed, data coding (floating point, integers, tensors, “spikes”, etc.), the architecture executing the code, and the technology in which the architecture realized are locally optimized by specialists that often don’t know the implications of their work on other domains.

**References**

[1] “António Guterres,” 10 9 2018. [Online]. Available: <https://www.un.org/sg/en/content/sg/statement/2018-09-10/secretary-generals-remarks-climate-change-delivered>.

[2] “Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts,” EUR-Lex, 21 April 2021. [Online]. Available: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206>. [Accessed 19 December 2023].

[3] “Proposal for a Regulation of the European Parliament and of the Council establishing a framework of measures for strengthening Europe’s semiconductor ecosystem (Chips Act),” 8 February 2022. [Online]. Available: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52022PC0046>. [Accessed 19 December 2023].

[4] “Information Communication Technologies (ICT),” [Online]. Available: [https://competition-policy.ec.europa.eu/sectors/ict\\_en](https://competition-policy.ec.europa.eu/sectors/ict_en). [Accessed 21 December 2022].

[5] “HiPEAC Vision,” HiPEAC, 17 January 2024. [Online]. Available: <http://vision.hipeac.net/>.

[6] “Computing Continuum website,” HiPEAC, [Online]. Available: <https://computing-continuum.eu/>. [Accessed 19 12 2023].

[7] “The TransContinuum Initiative,” ETP4HPC, [Online]. Available: <https://www.etp4hpc.eu/transcontinuum-initiative.html>. [Accessed 19 December 2022].

[8] “BLOOM: BigScience Large Open-science Open-access Multilingual Language Model,” HuggingFace, 6 July 2022. [Online]. Available: <https://huggingface.co/bigscience/bloom>. [Accessed 15 December 2023].

[9] “Snapdragon 8 Gen 3 Mobile Platform,” [Online]. Available: <https://www.qualcomm.com/products/mobile/snapdragon/smartphones/snapdragon-8-series-mobile-platforms/snapdragon-8-gen-3-mobile-platform>. [Accessed 14 December 2023].

[10] “MediaTek Dimensity 9300 deep dive: A true Snapdragon rival?,” 6 November 2023. [Online]. Available: <https://www.androidauthority.com/mediatek-dimensity-9300-explained-3381678/>. [Accessed 14 December 2023].

[11] “Digital Product Passport,” European Health and Digital Executive Agency (HaDEA), [Online]. Available: [https://hadea.ec.europa.eu/calls-proposals/digital-product-passport\\_en](https://hadea.ec.europa.eu/calls-proposals/digital-product-passport_en). [Accessed 15 12 2023].

# Acknowledgements

This document is based on the valuable input of HiPEAC members. The editorial board, composed of Marc Duranton (CEA), Koen De Bosschere (Ghent University), Paul Carpenter (BSC), Thomas Hoberg (Worldline), Charles Robinson (Thales), Tullio Vardanega (University of Padua) and Olivier Zendra (INRIA), would like to thank all the authors of the individual articles:

Jaume Abella (BSC), Rajendra Akerkar (Western Norway Research Institute), Djamila Aouada (Université du Luxembourg), Alessandra Bagnato (Softeam), Emmanuel Bertin (Orange and Telcom Sud Paris), Gaël Blondelle (Eclipse), Patrick Blouet (STMicro), Bart Coppens (UGENT) Adrian Cristal (BSC), Hugo Daniel Macedo (Aarhus University), Lieven Eeckhout (UGENT) Thomas Ernst (CEA), Safae Essafi Tremblay (VERSES.io), Christian Gamrat (CEA), Artur Garcia-Saez (BSC), Peter Gorm Larsen (Aarhus University), Katharina Grimme (IDC), Alexei Grinbaum (CEA), Miklós Györffi (Hungarian Research Network), Michael Henshaw (Loughborough University), Carles Hernandez Luz (Universitat Politècnica de València), Leonidas Kosmidis (BSC) Lennart M. Reimann (RWTH), Andreas Metzger (University of Duisburg-Essen), Miquel Moretó (BSC), Alix

Munier (LIP6) Claudio Pastrone (LINKS Foundation), Chase Pletts (VERSES.io) Peter Popov (City University London), Jérôme Quévremont (Thales), Lars-Åke Ragnarsson (imec), Jean-Pierre Raskin (Université Catholique de Louvain), Carlo Reita (CEA), Dan Richardson (Spatial Web Foundation), Walter Rudametkin (University of Rennes), Angela Salmeron (IDC), Claudio Sassanelli (Politecnico di Bari), Philippe Sayegh (VERSES.io) Dominik Sisejkovic (Bosch), Osman Unsal (BSC), Jack Vernon (IDC) Marcus Völp (Université du Luxembourg), Thorsten Weyer (Technical Hochschule Mittelhessen, Giessen), Lilia Zaourar (CEA).

The contents of the articles which have contributed to the HiPEAC Vision 2024 have also benefited from indirect contributions from a large number of individuals, too many to enumerate, via accidental conversations, exchange of opinions, on-the-fly comments, etc.

The editorial board would also like to thank Eneko Illarramendi, Vicky Wandels (both Ghent University) and Madeleine Gray for their support and contributions. Special thanks to the Belgian comic artist Arnulf who provided a light-hearted, often satirical take on the content of this Vision document.

# Process

The HiPEAC Vision presents the trends that have an impact on the high performance, edge and cloud computing and related communities. It consists of a standalone vision text and a number of independent articles that underpin its recommendations. The articles and recommendations are based on information collected through a number of channels.

- Meetings with teachers and industrial partners at the ACACES 2023 Summer School;
- A survey circulated to all HiPEAC members, which received 29 responses;
- Several consultation meetings:
  - 30 May 2023: Consultation meeting on new software paradigms,
  - 31 May 2023: Consultation meeting on design tools for software and hardware,
  - 20 June 2023: Consultation meeting on guardian angels and the next web,
  - 22 June 2023: Consultation meeting on cryptocurrencies, blockchain and DLT,
- 10 August 2023 (online): consultation meeting on sustainability.
- Five in-person editorial board meetings and 9 by Webex;
- Exchanges with other organizations, such as ETP4HPC, ECSO, KDT, AIOT, ADRA, BDVA, SNS, NESSI, FIWARE, Destination Earth, Eclipse foundation, etc...);
- Feedback from presentation of the HiPEAC Vision 2023 at several conferences and workshops and from exchanges with the DG CNECT;
- Feedback from a presentation of an early version of the HiPEAC Vision 2024 at CONNECT University (viewed by 245 people) on 10 November 2023.

Note: this document is called a “Vision” because it is the result of the interpretation of the trends and directions as seen by the HiPEAC community. As HiPEAC has no direct power to enforce the recommendations, the timeline associated with the potential implementation of the recommendations is uncertain; this is why the document is not a roadmap per se.



# HiPEAC Vision 2024

The HiPEAC Vision 2024 recommendations focus on adapting European computing systems to rapidly evolving technologies and intense geopolitical and environmental pressures. Central to this is the development of the “next computing paradigm” (NCP), which integrates various key technologies – including artificial intelligence, (AI), the internet of things (IoT), and digital twins into a federated ecosystem. This paradigm emphasizes spatial computing, AI at the edge, dynamic web integration, and efficient orchestration of distributed services.

Advancing the NCP involves developing technologies essential for the NCP, making Europe a leader in AI, particularly for widespread edge use, innovating in hardware solutions, prioritizing cybersecurity, and promoting sustainability assessment for all systems. Global thinking and cross-domain collaborations are crucial.

## Recommendations



### 1. Operationalize the NCP:

- Develop 4D-aware implementation technologies, standardizing representation and protocols for encoding physical objects and spaces, supporting mobile computation, and powering 4D-enabled operations.
- Augment APIs for interoperability, enhancing APIs with specifications for non-functional properties and dynamic service composition.
- Enable the mobility of computation, relocating data and processes as needed.
- Adopt (generative) AI at the edge for greater efficiency and privacy, and reduced latency.
- Develop AI-powered edge orchestrators that can dynamically combine services based on user needs.
- Encourage non-proprietary integration by supporting open standards and platforms.
- Initiate proof-of-concept efforts, creating demonstrators to showcase the advantages of NCP technologies.



### 2. Make Europe a leader in controlled AI development:

- Support EU growth in AI domains by investing in AI research and infrastructure.
- Develop foundation models based on “European” values and reflecting regional needs.
- Promote open-source AI models to enable access to shared AI resources.
- Develop local AI solutions and specialized accelerators for edge devices.
- Use AI for software and hardware development, upskilling engineers and researchers.
- Develop policies around AI accessibility and societal impact, ensuring equitable benefits from AI.
- Ensure “correctness by construction”, automating AI output verification for trustworthiness.



### 3. Innovations in hardware:

- Continue to improve performance and energy efficiency, including by exploring and integrating new technologies.
- Promote interdisciplinary research, driving new hardware paradigms.

- Explore innovative architectures for data-intensive computing.
- Develop a full European ecosystem, promoting fast prototyping and specialized architectures and developing the chiplet/interposer European ecosystem.
- Ensure sustainable hardware development, focusing on energy and emission reduction, and recyclability.



### 4. Cybersecurity as a core aspect:

- Address vulnerabilities, building systems with early consideration of cybersecurity and privacy.
- Reduce dependence on external ICT, using trusted European companies or open-source solutions instead.
- Address security challenges in large language models (LLMs).



### 5. Sustainability in computing:

- Create validated lifecycle models that model environmental impact.
- Develop sustainability-focused design, accounting for environmental costs in product design.
- Create viable sustainable business models for the ICT industry.
- Create ICT solutions for green applications in other industrial sectors.



### 6. Think globally:

- Promote collaboration: encourage teamwork across European research and technology groups.
- Cross-domain project calls: foster interdisciplinary research for innovative solutions.
- Competence centres: create centralized European expertise hubs.
- Build open-source ecosystems to accelerate innovation and accessibility.
- Multi-dimensional tooling: develop tools supporting capacity to address complex and critical challenges.
- Adopt a holistic approach to efficiency, emphasizing global co-design and system thinking.

