Recommendations

State of the (European) Union

Build science and technology clusters

According to the Draghi report, (i) the EU has only one science and technology (S&T) cluster (ranked 12) in the global ranking of the 20 largest S&T clusters of the world, and (ii) European companies have difficulties scaling up from startup to unicorn and beyond. Science and technology clusters are ecosystems that help new technology companies to hatch and grow by providing world-class research facilities, the proximity of a world-class higher education institution providing a talent pool, incubators and accelerators, growth capital, a favourable legislative framework, and first and foremost a vibrant community of entrepreneurs. Many of the global technology companies grew from such a cluster, and the fact that Europe has only one such cluster in the top 20 is problematic. Europe should therefore actively promote the creation of European S&T clusters in major urban areas and help them grow to a scale that they can support scaleup companies.

Introduce ARPA model of challenges

ARPA (Advanced Research Projects Agency) in the US funds high-risk, high-rewards projects to generate transformative technologies. ARPA focuses on radical innovation and is willing to accept failure as part of exploring new ideas. Projects are quite short (two to five years) and must show measurable progress quickly. They are led by entrepreneurial programme managers who have a vision for technology breakthroughs, scout for innovative ideas, assemble the best teams and take corrective action if milestones are not met (including termination). This introduces a new R&D culture: fast, milestone-based, competitive, risk-tolerant, visionary, agile. Europe should use a similar model to tackle some of the grand challenges.

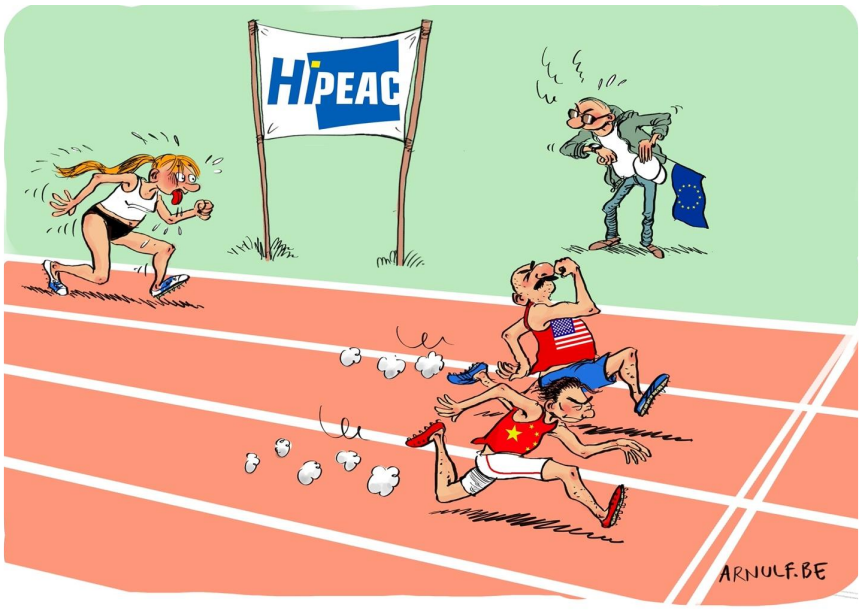
Stimulate pre-competitive procurement

A weakness of the current publicly funded research programmes in Europe is a failure to realize the full commercialization potential of research results. In many cases, the research results could be a good starting point for a spin-off company, but if nobody involved in the project has the ambition to start a company, the results are not commercially exploited. The reasons are well known: the principal investigators have a stable position in a research institute or company, and are not looking for an entrepreneurial adventure, and the goal of the PhD-students is to finish their PhD, not to create a company. Another barrier is that the gap between a proof of concept and a product is large, and researchers seldom have the business skills to close that gap.

Pre-competitive procurement follows a different approach. A government orders an innovative product or service that does not yet exist and creates a tender for a company or consortium of companies and universities / research and technology organizations (RTOs) to develop it.

This is how the world got COVID-19 vaccines: the first promising results of the phase I human trial were announced on 18 May 2020, that is, 137 days after the identification of the virus. The company was Moderna, a company only founded in 2010 in a sector where it is very difficult to bring a product to the market because it has to be clinically tested and approved by governments. Less than one year after the identification of the virus, the vaccination campaign was already rolled out at globally – this is the typical time between launching a call for research projects and the kick-off of the first projects.

Pre-competitive procurement not only shortens the execution time of the projects, but it also increases the likelihood of commercialization because delivering a working product or service is the task given to the consortium.



Next Computing Paradigm

Create digital envelopes

Create a **"digital envelope"** for integrating "anything" (person, company, physical entity, computing device) in a digital space allowing access to multiplicity of services – i.e. building on the concept of "anything-as-a-service" (XaaS) – in an interoperable way. This digital envelope would allow live migration of compute components and a **runtime evolving infrastructure**to support **deployment on a continuum**ranging from resource-constrained edge devices to data centres, allowing for dynamic resource pooling and efficient sandboxed execution of collaborative, migratory compute components offering services.

1. An intelligent **digital agent** able to pursue goals legitimately assigned to it. The intelligent digital agent would be capable of direct execution as well as of **orchestration**. The former would be required when seeking the set goal required local actuation, the latter when the task resulting from the set goal required remote execution.
2. **Sensors**, to pull digitalized inputs from designated sources (in the physical world or other digital envelopes).
3. **Actuators**, to push computed outputs into designated targets (physical things or digital envelopes). The fabric resulting from interconnecting digital envelopes that

provide and require services from one another to pursue assigned goals will operate in genuine XaaS modality.

> "Digital enveloping" is the technology-enabled phenomenon by which any item of reality – human, material or immaterial – can be associated with a computable digital representation capable of delegated autonomous action. The notion of delegated autonomous action entails two fundamental traits: that of delegation, which suggests a higher (human) authority that requires some action to be taken (in part) in the digital space; and that of autonomy, which suggests that the pursuit and execution of the required action is carried out by autonomous executable agents that operate within the remits of delegated authority.
>
> The capacity for delegated autonomous action is provided to individual digital envelopes by the combined operation of three key components:

These solutions enable the live migration of compute components across the edge-to-cloud continuum – therefore services – ensuring continuity while addressing latency, privacy, security, risk management, validation mechanisms and context requirements. This is essential to optimize user and infrastructure needs dynamically. In relation to real-world services or actions triggered by a digital envelope, location and time identifiers are assigned, and inter-envelope mechanisms support local vs global optimizations including for safe interactions, managing complex interdependencies and conflict resolution. The next computing (NCP) exemplifies the notion of continuum. Agreed standards are key for interoperability.

AI-powered orchestrators

Artificial intelligence (AI)-powered orchestrators will be an essential capability of the intelligent digital agent of the digital envelope. AI-powered orchestrators will be developed for the edge – which is strategic as it is located the nearest to the final user – in a manner that can dynamically combine collaborative compute components into executable applications tailored around specific user needs. The task of the orchestrators is to decompose goals set by the user (in the broader term, including human user, company or another permissioned orchestration) into a set of services that cooperate to achieve the set goals. These orchestrators could be themselves generated by (federated) **generative AI (genAI) engines** (supported by more classical algorithmic approaches) located at the edge and capable of collaboration with other orchestrators within federated zones.

Space- and time-aware protocols

Expand and adapt web-level protocols and associated standards by enhancing the existing suite of HTTP-based protocols to be both spatially aware and time-sensitive. This will allow web-level interactions between NCP's migratory compute components to account for 3D physical space and real-time communication, drawing on technologies like WebRTC to manage time-sensitive tasks effectively and the spatial web (IEEE P2874, OpenUSD, …).

Interoperable contract-based API specifications

Establish interoperable, contract-based application programming interface (API) specifications – usable by expanded web-level protocols – ensuring that interconnected services communicate with clear expectations of both functional and non-functional performance. These contracts, similar to service-level agreements (SLAs), should detail the conditions under which services will optimally perform, including non-functional requirements, ensuring smooth integration and reliable service delivery within the NCP framework. These APIs should account for non-functional properties like latency, cost, and

performance. The resulting model should ensure that an API not only promises to deliver a service but also specifies the conditions under which it can perform optimally. The API should also be compliant with the currently proposed API for large language models LLMs `[OpenAIFunction]` [BerkeleyFunction].

Promoting these standards in relevant standardization bodies is essential for fostering interoperability and consolidating development conditions through standardized benchmarks, testing methodologies, and best practices. This will ensure that implementations can be effectively and securely integrated, improving overall system efficiency and reliability, and enabling the creation of an interoperable business ecosystem of services and orchestrators.

Artificial Intelligence

Develop distributed agentic AI (specialized action models)

The **development of specialized action models (SAMs)** acting as service is important and can be developed in Europe. These SAMs, small and specialized models that can interact with their environment, should operate in a distributed infrastructure and an ecosystem should be created to support research, development and business around them. These models need to be refined, optimized, and reduced in size to improve efficiency. These SAMs can be optimized from more general foundation models by an ecosystem of companies providing their optimized SAMs in a marketplace so that they can be dynamically discovered and used by the orchestrators.

Develop orchestrating technologies for distributed agentic AI, blueprint for NCP orchestrators

We call agentic AI a set of specialized AI agents working together to accomplish a common goal. An AI agent is synonymous with an SAM in this discussion: an AI that can perceive and act, having impact on the virtual or real world. **The orchestration technologies** should take into account all the requirements, that can select the best SAMs for the required tasks and dynamically activate them. The first steps could be very agentic-AI-centric (relying on already

existing technologies used for orchestrating AI agents), but they should be blueprint and evolve towards an orchestration system for the NCP. These orchestrators must be developed for the edge – or near the final user – and dynamically combine SAMs into executing personalized applications in response to user needs.

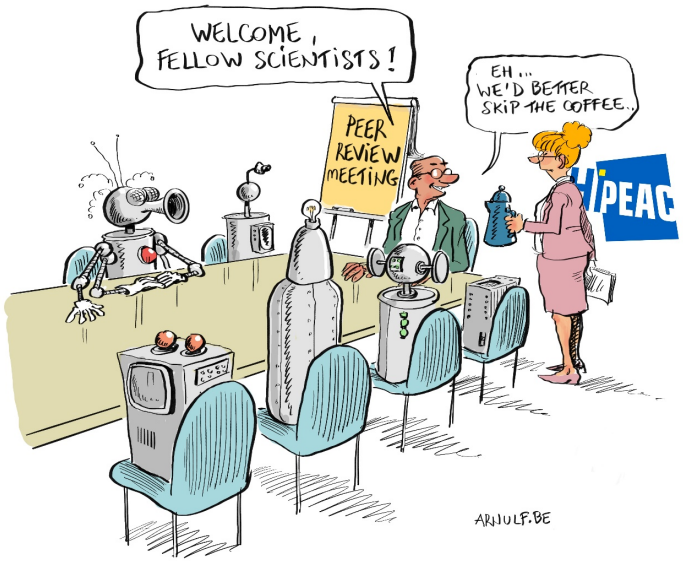
Establish open protocols for these "distributed agentic AI" systems to facilitate seamless interaction among distributed AIs from different origins

Protocols and specifications that group all requirements, existing ideas and proposals together in a single consortium to develop an open source "de facto" (before official standardization) standard protocol that takes into account all the good ideas of various researchers and organizations, so that it will be sound, future-proof, recognized and accepted. The requirements are:

1. It does not solely rely on functional requirements (e.g. the textual representation of prompts and responses).

1. It also incorporates non-functional requirements (providing sufficient information for the orchestrator to select the appropriate services, such as based on criteria like response time, potential level of hallucinations, environmental impact, cost, localization, privacy of data, etc.).

The recommendation to develop generative AI at the edge (AI) is still important, but it is more in development and implementation mode now (for example, in Apple intelligence). We should continue developing solutions that allow embedding generative AI at the edge in order that **human users can be provided with natural interfaces** (voice, gesture, eye movements, touch) to the digital world, with more energy efficiency, reduced latency, lesser communication overhead, and greater privacy. This is important to reduce the difficulties to access the digital world and decrease digital illiteracy.



New hardware

Specialized hardware (HW)

The development of efficient hardware is essential for running services, orchestrators and SAMs efficiently at the edge and within federated networks. Europe must address memory costs (for AI), energy consumption, and ecological impact, potentially leveraging non-volatile memory for direct edge execution. Additionally, the next generation of SAMs should incorporate learning through experiences or allow to the efficient execution of digital twins to maintain Europe's competitive edge in AI (embedded AI). In the field of AI accelerators, the focus should be on inference (becoming more and more important with the approach pioneered by OpenAI o1 and o3) or on fine tuning. Reducing the transfer of data is key to reach lower levels of power consumption. This can be achieved with near- or in-memory computing (NMC or IMC), direct execution from the storage of parameters (hence eliminating the need for RAM), etc…

Beyond purely digital hardware (HW)

Investigation of new **accelerators using non digital technologies**, going from exact computations (digital computation) to more approximate computing (neural networks are universal approximators, quantum computing results are stochastics, optimization techniques using Bayesian, Ising approaches can solve complex problems) should be also investigated in the context of providing more efficient services to the next computing paradigm (NCP) ecosystem.

Tools

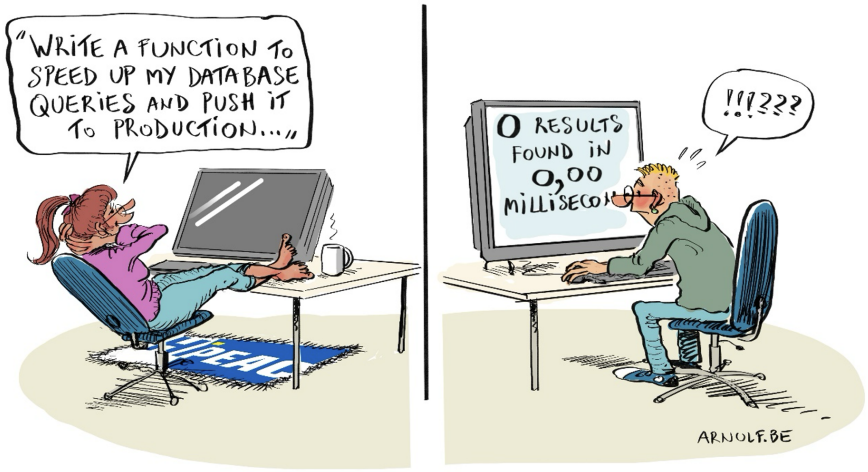
Promote the use of AI in software development

Research, prototype and deploy AI-assisted software development environments, while implementing robust measures to ensure correctness, safety, security, confidentiality, and regulatory compliance. This will help balance the rapid adoption of AI with the need for

secure and reliable systems. It should also help non specialists to be able to create efficient software and increase the productivity of developers.

Promote the use of AI in hardware development

Research, prototype and deploy open AI assistants for hardware development, increasing the productivity for designing new, efficient hardware and decreasing the time to market. This is a key element for Europe to stay in the hardware race. The use of AI should be a collaboration between humans and AI systems, as promoted in previous HiPEAC vision as 'centaur' teams. The focus should be on domains that are still open, like architecture search and exploration, rather than on optimizing the floor-planning, which is already covered by various companies.



Cyber-Physical Systems

Accelerate cross-disciplinary joint research

The technology domains contributing to Cyber-Physical Systems research call for investment in tools, methods and cross-technology community initiatives to tackle the multi-stakeholder research barrier - especially arising for a technology bridging diverse complex knowledge domains and applied at higher levels of a system where there are many more interactions with the technology to consider - higher-order integrated research. This will accelerate progress towards the Next Computing Paradigm and CPS research as well as technology infrastructure updates by tackling the challenges of diverse knowledge domain perspectives and enabling access to the bigger picture. In particular: 1) A new R&D dimension to really boost our capability for highly complex and cross-domain integrated research activities. Just as we have different approaches for building windows and houses, there is need to establish tools and methods supporting higher order integrated research. This is especially a case in point for the highest integration levels of CPS research where most impact and value generation can be expected. Adapted or new tools and methods for convergence, with strong public engagement, should support terminologies (e.g. wiki-style trusted glossary), concept sharing (e.g. modelling), knowledge sharing (e.g. ontologies via Protégé), consistent evaluation approaches and global visualisations, including non-technical domains. 2) Existing communities should establish a centralised CPS association to unify efforts, promote knowledge exchange, and align standards; 3) Additionally, frameworks for integrating AI/ML into CPS must address safety, security, and ethics,

ensuring dependable systems for sectors like healthcare and transport. These actions are vital to Europe's sovereignty and global leadership in CPS advancements.

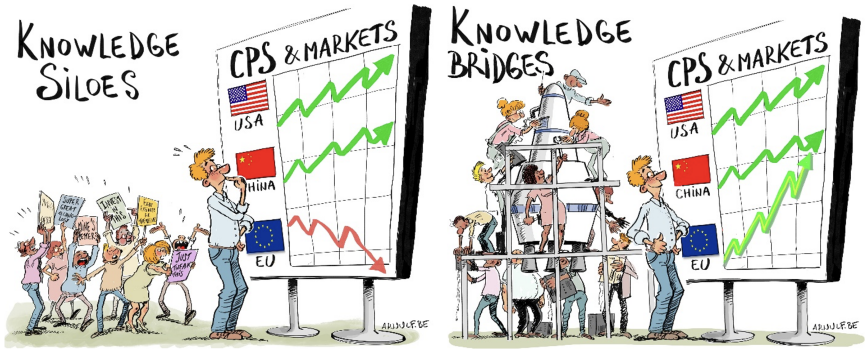
Redefining dependability for CPS adaptability and technology integrations

CPS depend on safety, security, and performance properties to govern what they can achieve and qualify technologies for use. CPS contributing communities encourage: 1) Solutions to migrate from legacy approaches that minimise interactions of these properties to instead maximised interactions for optimum system adaptability. These properties impose constraints on available choices we have at design and in operations, which are compounded by ruling out choices where trade-offs would be required. Techniques such as combined analysis, evaluation and knock-on effects should be advanced for handling these properties. Establishing an approach, considering tools and methods referring to best practice, is needed to account for the interdisciplinary integration overheads between these traditionally distant domains, but also with the rest of the system. This is crucial in CPS for enhancing scope of AI/ML and IoT usage, as well as other technologies. 2) A new way of thinking is needed for treating interconnected systems with CPS - dependability considered in a modular fashion - with hazard analysis techniques likes STPA extended, including for man-machine teaming and AI complexities. We encourage also frameworks for risk assessment in relation to AI/ML to be established and considering adaptive risk management strategies in the context of these interconnected critical systems. This moves forward with trustworthy CPS in sectors like AI-enabled autonomous systems.

AI-performance-defence guarantees for real-time interconnected systems

Future CPS require advanced technologies to address challenges in performance characterization, damage containment, and operational feedback. CPS contributing communities encourage: 1) Real-time methods ensuring deterministic multi-tasking environments and verifiable AI/ML performance. In complement, there should be an extension of defence mechanisms and feedback loops, which is essential for preventing damage propagation and enabling iterative improvement. Solutions should emphasize distributed architectures, particularly edge computing, and include digital twin capabilities for predictive insights. 2) Comprehensive uncertainty quantification, real-time monitoring, run-time verification, and data flow tracking will enhance trustworthiness. These advancements will support supervisory control and ensure dependable CPS operations, even in rapidly evolving and uncertain environments like AI-enabled applications.

These three recommendations are detailed next. Due to the multi-domain nature of CPS research they have also been extended as an associated white paper [1].



Cybersecurity

Software supply-chain cybersecurity

Reinforcing software supply-chain cybersecurity is crucial given the wide impact of attacks spread through the supply chain, which is all the more important given the large number of components in the next computing paradigm (NCP). Develop code and component analysis technologies for cybersecurity that scale up and support trusted orchestrators, services and communications.

Comprehensive safety, security, and performance coupling requires standardized software vulnerability representation. Increased interconnectivity requires new technologies to isolate threats and proactive cyber-risk management. Develop secure software package and service management that balances usability with strong security.

AI for cybersecurity

To enhance NCP cybersecurity in a scalable way, develop i) advanced artificial intelligence (AI) models, including large language models (LLMs), for threat detection and ii) autonomous systems for mitigation (e.g. isolating compromised NCP components, patching vulnerabilities, or restoring services). Utilize federated AI for its decentralized, privacy-preserving and scalable models in the NCP massively interconnected context. Rely on EU-based open AI models and datasets to strengthen EU cybersecurity, sovereignty, and competitiveness.

Reinforced cybersecurity of AI

Secure AI training methodologies and validation procedures, as well as adversarial defences, are needed. LLM prompt injection attacks must be a major concern, addressed by the development of tools to detect and secure against these, and by establishing benchmarks for prompt injection prevention and response. AI security standards should be established by developing certification procedures to guarantee that LLMs and AI systems adhere to stringent security standard, possibly requiring security audits for AI systems. These efforts should rely on EU-based open AI models.

Sustainability

Validated life-cycle models for computing

The information technology (IT) community should further develop validated life-cycle models for its own products and services. These models should comprehensively account for the total environmental impact of the production and disposal of the product, commonly known as embodied emissions. This includes the impact of mining, water usage, the use of chemicals in production, and end-of-life processing.

In addition, the model should also estimate operational emissions. This information should be included in a digital product passport (DPP) containing information about the environmental impact comparable with the information on pre-packaged food products or power-efficiency information on household appliances. This information will help consumers to make informed choices about sustainability. The digital envelope of a device should be able to return this information to e.g. an orchestrator to enable it to select the services that optimize the sustainability requirements specified by the owner of the orchestrator.

Sustainability-focused design methodologies and business models

Detailed life-cycle models will help designers make the most effective eco-design decisions. To be effective, design tools should automatically include the environmental impact of the components and technologies used in the design, without putting the burden on the designer. Incorporating repairability, reusability, recyclability, and end-of-life processing considerations from the beginning of the product development process will also lower the environmental impact of the final design.

Inevitably, reducing the environmental impact of a product will have an impact on companies' business models. Designing products that last longer will reduce sales of new products and hence lower the profitability of the company. This can only be mitigated by developing new business models, based on extra services: maintenance, repair, disposal, … up to completely replacing the ownership of hardware by a service contract. The goal should be to bring services to the market with the least environmental impact possible (which in practice means with the least amount of hardware, and the lowest power consumption).

References

`OpenAIFunction:` OpenAI Platform: Function Calling. https://platform.openai.com/docs/guides/function-calling

`1:` Charles R. Robinson et al. (2025). Extended Recommendations for Advances on Cyber-Physical Systems. Zenodo. https://doi.org/10.5281/zenodo.14624958