



RESEARCH & INNOVATION ROADMAP

Cognitive Computing Continuum

WHITE PAPER

Second Version - March 2025

www.eucloudedgeiot.eu

Executive summary

This Research & Innovation Roadmap, developed under the NexusForum.EU project and supported by Horizon Europe, outlines a research and innovation trajectory towards achieving a federated European multi-provider, AI-driven computing continuum, to enable and support data-driven innovation and AI deployments in Europe. It complements the technology developments of the European Alliance for Industrial Data, Edge and Cloud and the new IPCEI-CIS, providing a long-term vision bridging industry needs with excellent research in cloud, edge, and AI technologies. The roadmap also addresses the current EU policy and legislative framework, the implications of which for future technological developments are further analyzed in NexusForum policy reports.¹

This is the second public version of the Cognitive Computing Continuum Research and Innovation Roadmap for Europe. A final version will be released close to the end of the NexusForum project in mid-2026, and will further identify strategic areas for collaboration with South Korea and Japan. NexusForum will implement a public consultation process to gather and incorporate feedback from the broader community and stakeholders in cloud, edge, AI, and industry sectors that could benefit from these technologies. Relevant experts and stakeholders are welcome to actively participate in the NexusForum working groups, where the roadmap will be further discussed and elaborated.

Towards a European Cognitive Computing Continuum

The European Union stands at a critical juncture in shaping its digital future. The importance of cloud and edge computing, together with artificial intelligence (AI), is well-recognised as a strategic enabler for innovation and competitiveness in an increasingly digital and data-driven society. Despite this strategic importance, the market for cloud services and AI is currently dominated by a few mostly non-European players, and the European cloud market remains highly fragmented.

There is a need for organisations and citizens in the EU to consolidate their IT systems and data, and gain control over how they are collected, stored, and used, in line with European rules. The European data strategy aims to address this need and establish a single unified (European) market for data, where data can move more freely and the benefits of them can be better shared. This unified single market will hopefully reduce vendor lock-in and data silos, and ultimately lower the barriers for data-driven innovation across sectors.

To realize a European Cognitive Computing Continuum, the EU must overcome market fragmentation, foster interoperability, and ensure compliance with European regulations, values, and ethical AI principles.

A Strategic Vision for European Digital Sovereignty

The European Cognitive Computing Continuum should be an interconnected, federated ecosystem in which European cloud and edge service providers collaborate to deliver secure,

¹ https://eucloudedgeiot.eu/wp-content/uploads/2024/10/D3.1%E2%80%93DigitalPolicyReport_-_Final.pdf

high-performance, and energy-efficient data processing capabilities, with the possibility of provision from suitable other jurisdictions such as Japan and Republic of Korea.

This roadmap offers an in-depth overview of the technology areas and associated subjects aimed at developing a competitive, secure, energy-efficient, climate-neutral, and AI-powered Cognitive Computing Continuum.

The current version focuses on addressing the key priorities of recent EU policy initiatives (Competitiveness Compass,² Connected Collaborative Networks³), on responding to the current geopolitical landscape and market conditions (A competitiveness strategy for Europe,⁴ Much more than a market⁵), on the regulatory framework (NIS2 Directive,⁶ Cyber Resilience Act,⁷ Chips Act,⁸ EUCS⁹) and on societal needs in Europe, pursuing mainly:

- The enhancement of European competitiveness in the Edge and Cloud domain to balance the current market dominance of the three non-European hyperscalers, while balancing sovereign cloud initiatives with access to essential technologies.
- EU digital sovereignty in the Computing Continuum domain, where technology sovereignty is essential for security and regulatory compliance.
- The generative AI race and meeting computing requirements to support the future needs of foundational AI models.
- Achieving carbon neutrality and ensuring energy-efficient computing resources and software.
- Incorporating perspectives from international partner countries with strong positions in the cloud computing continuum.

Key Recommendations

This roadmap highlights the following strategic research and innovation priorities to realise the vision of a European Cognitive Computing Continuum.

1. Create the Foundations for a Secure, Interoperable, and Sovereign European Computing Continuum

- Adopt a holistic and systems-centric approach to cybersecurity, analysing cascading risks and systematic risks in the European Computing Continuum.

² https://ec.europa.eu/commission/presscorner/detail/en/ip_25_339

³ <https://digital-strategy.ec.europa.eu/en/library/white-paper-how-master-europes-digital-infrastructure-needs>

⁴ https://commission.europa.eu/topics/eu-competitiveness/draghi-report_en#paragraph_47059

⁵ <https://www.consilium.europa.eu/media/ny3j24sm/much-more-than-a-market-report-by-enrico-letta.pdf>

⁶ <https://digital-strategy.ec.europa.eu/en/policies/nis2-directive>

⁷ <https://digital-strategy.ec.europa.eu/en/policies/cyber-resilience-act>

⁸ https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-chips-act_en

⁹ <https://www.enisa.europa.eu/publications/eucs-cloud-service-scheme>

- Harmonise the markets and regulations within Europe, and support the development of ecosystems and mechanisms for interoperability, to ensure seamless service integration and data portability.
- Strengthen governance and compliance mechanisms aligned with European regulations (NIS2, Cyber Resilience Act, EUCS).
- Consider open-source hardware and software technologies as a geopolitical concern, and ensure transparency and long-term sustainability of European digital infrastructure by supporting open-source governance models, for example based on RISC-V.
- As the supply of European semiconductor technologies is strengthened, develop an early focus for their uptake by creating a strong demand and market for them in advance.

2. Invest in Intelligent, Autonomous, and Resilient Management of a Multi-provider Computing Continuum

- Develop AI-driven orchestration frameworks that optimize workload distribution across federated cloud-edge infrastructures.
- Advance federated computation models to support privacy-preserving AI, ensuring data security across multiple providers.
- Enable hyper-decentralized architectures that enhance resilience, scalability, and self-management of computing resources.
- Invest in confidential computing and homomorphic encryption techniques to secure data and computations.
- Support AI-powered predictive maintenance, debugging, and self-healing cloud-edge infrastructure to enhance reliability and reduce operational costs.

3. Enable Data-driven Innovation, and the Development and Deployment of AI in the Computing Continuum

- Expand the development of **Common European Data Spaces** to facilitate cross-border data sharing in compliance with GDPR, the AI Act, and the Data Act.
- Enable **portable AI applications and services** that can run seamlessly across different cloud-edge providers, including EuroHPC infrastructure.
- Develop **middleware solutions** and compiler technologies to improve portability, interoperability, and security in heterogeneous multi-provider cloud-edge systems.
- Advance research in **federated training and deployment of generative AI models** across distributed datasets.
- Explore integration of **quantum and neuromorphic computing capabilities** to drive future AI acceleration.

4. Create a Sustainable and Energy-Efficient European Computing Continuum

- Develop and deploy carbon-aware computing strategies to reduce the environmental impact of data centers and AI workloads.

- Leverage AI for optimizing holistic energy consumption across cloud-edge infrastructures, including workload balancing with data center operations.
- Invest in middleware solutions and co-design of software and hardware to optimize code execution and reduce energy use of AI-specialised computing platforms.
- Promote the development of new cooling technologies and **waste heat recovery systems**, and integrate renewable energy into cloud and edge data centers.

5. Converging Telecommunications, Computing, and AI Infrastructure

- Invest in an **Open Radio Access Network (O-RAN)** and next-generation telco-cloud architectures to create a flexible, open, and interoperable network ecosystem.
- Enhance **seamless data connectivity** by improving predictive handover and intelligent network management across different networks.
- Integrate European **high-performance computing (HPC)** to enable advanced AI and scientific computing workloads.
- Integration of **Information Technologies (IT) and Operational Technologies (OT)** to drive digital transformation in industrial sectors.

Key priorities include the acceleration of sector-specific digital transformation through the development, and implementation of various next-generation technologies, such as: the AI-powered computing in automotive for connected and autonomous vehicles; AR/VR technologies to enable immersive industrial, healthcare, and training applications; cyber-physical convergence through digital twins and smart environments for real-time decision-making; Industry 4.0-ready cloud-edge solutions that support smart manufacturing, logistics, and critical infrastructure.

Conclusions

To ensure Europe's digital sovereignty and competitiveness in the global AI and cloud landscape, it is critical to accelerate strategic investments and collaboration within Europe and with suitable external partner countries, to align key digital technological advancements with European values. It is also critical to develop in Europe strong demand alongside strategic investments in strengthening supply of strategic technologies.

Table of contents

| | | |
|----------|----------------------------------------------------------------------------------------------|-----------|
| 1 | Introduction | 9 |
| 1.1 | Aims and overview of this document | 10 |
| 2 | A European Perspective: Markets, Laws, and Initiatives | 14 |
| 2.1 | European market context..... | 15 |
| 2.2 | European legal context | 17 |
| 2.3 | European initiatives related to cloud and edge | 19 |
| 3 | Towards a European Cognitive Computing Continuum..... | 22 |
| 3.1 | Fragmented European data centre market..... | 23 |
| 3.2 | Fragmented European cloud market | 25 |
| 3.3 | Convergence of networking and compute, and open radio access networks | 26 |
| 3.4 | Convergence of Operational Technologies and Information Technologies | 27 |
| 3.5 | Establishing a European ecosystem based on RISC-V | 28 |
| 3.6 | EuroHPC and the establishment of AI Factories | 30 |
| 3.7 | Access to European data, common data spaces, and data privacy | 31 |
| 4 | Transversal topics | 33 |
| 4.1 | Digital sovereignty..... | 33 |
| 4.2 | Open Source | 34 |
| 4.3 | Sustainability and Energy Efficiency..... | 35 |
| 4.4 | Cybersecurity | 37 |
| 4.5 | Interoperability | 40 |
| 5 | AI for Cloud-Edge: Orchestration and Managing a Multi-Provider Continuum..... | 43 |
| 5.1 | Sustainable, energy-efficient, and energy-grid-aware Computing Continuum..... | 43 |
| 5.2 | AI-assisted operation maintenance of large-scale systems | 45 |
| 5.3 | New distributed data processing paradigms for the Computing Continuum | 47 |
| 5.4 | Safety-critical applications in the Computing Continuum | 48 |
| 5.5 | The continuum performance: cross-layer optimisation..... | 50 |
| 5.6 | Towards a hyper-decentralized Computing Continuum | 51 |
| 6 | Cloud-Edge for AI: Enabling and Facilitating AI Applications Across the Continuum. 53 | |
| 6.1 | Portable AI applications and an open AI ecosystem | 54 |
| 6.2 | Integrating HPC in the Cognitive Computing Continuum | 55 |
| 6.3 | RISC-V and new processing initiatives in the Cognitive Computing Continuum | 57 |
| 6.4 | A middleware toolkit for the Computing Continuum | 59 |
| 6.5 | Operationalization of future AI systems in the Computing Continuum (GenAIOps) | 60 |

| | | |
|-----------|--------------------------------------------------------------------------------------------|-----------|
| 6.6 | Federated computation for Foundation Models..... | 61 |
| 6.7 | Large-scale testbeds for AI services in the Computing Continuum | 63 |
| 6.8 | Data privacy and security in AI services..... | 64 |
| 6.9 | Confidential computing | 65 |
| 6.10 | Data spaces for AI in the Computing Continuum | 66 |
| 7 | Telco Cloud-Edge: Telco as One of the Main Tenants and Infrastructure Providers ... | 69 |
| 7.1 | Open radio access network (Open RAN) | 71 |
| 7.2 | Seamless data connectivity and predictive handover across different networks | 73 |
| 7.3 | Generative telco cloud | 75 |
| 7.4 | Resilient on-prem 5G/6G edge-cloud for Industry 4.0 and beyond..... | 77 |
| 8 | Digitalisation of Industry Sectors: Requirements from Next-Gen Applications | 79 |
| 8.1 | Automotive and the continuum | 79 |
| 8.2 | Augmented Reality/Virtual Reality technologies..... | 80 |
| 8.3 | Cyberspace and physical space fusion | 81 |
| 9 | Disruptive impact initiatives | 83 |
| 9.1 | Neuromorphic systems | 83 |
| 9.2 | Space Edge | 84 |
| 9.3 | Integration of quantum computing infrastructure | 85 |
| 9.4 | Quantum and classical computing fusion | 87 |
| 10 | Conclusions..... | 88 |

List of figures

| | |
|----------------------------------------------------------------------------------------------|----|
| Figure 1 Illustration of research & innovation subtopics | 13 |
| Figure 2 Productive investment over two decades, EIB 2024 | 16 |
| Figure 3. EU laws over time and their impact in the European Computing Continuum..... | 17 |
| Figure 4 How European initiatives and policies related to cloud computing fit together. | 20 |
| Figure 5 Main dimensions of the emerging Cloud-Edge Computing Continuum infrastructure. ... | 23 |
| Figure 6 Large Language Models Operations (LLMOps) | 62 |

1 Introduction

Cloud and edge technologies are key strategic enablers for innovations and uptake of emerging technologies, such as artificial intelligence, the Internet of Things (IoT), and 5G. As Europe positions itself as a leader in a data-driven society and development of trustworthy AI, the European Union acknowledges the significance of cloud and edge computing in shaping its digital future. This significance is reflected, for example, in the digital targets for 2030 of Europe's Digital Decade policy programme.¹⁰ The European data strategy further underscores this significance and the importance of *“investing in next generation tools and infrastructures to store and process data”* and *“joining forces in European cloud capacity”* to create a single market for data in Europe based on European rules, values and regulations, in particular privacy and data protection.¹¹ These technologies are key building blocks and enablers of the European *common and interoperable data spaces* meant to overcome legal and technical barriers to data sharing to enable data-driven innovation in Europe at scale.¹²

To address the urgent need to *“support research, development, and first industrial deployment of advanced cloud and edge computing technologies across multiple providers in Europe”*, the Commission approved up to €1.2 billion of state aid for the IPCEI Next Generation Cloud Infrastructure and Services (IPCEI CIS) project.

Similarly, the European Alliance for Industrial Data, Edge and Cloud was set up to strengthen Europe's leadership position on industrial data and support the development of next-generation cloud and edge technologies. It brings together businesses, Member State representatives, and other relevant experts, facilitated by DG CONNECT, aiming to *“strengthen the position of EU industry on cloud and edge technologies”*.¹³ The Alliance provides recommendations and investment roadmaps for the development of European cloud and edge.

This critical role of computing technologies for the future of European competitiveness and digitalisation of critical sectors is again stressed in the recent Draghi report.¹⁴ Under the topic of *digitalisation and advanced technologies*, Part B of the report identifies and recommends the following three strategic priorities for the coming decade, cited from the report:

- **High-speed/capacity broadband networks and related equipment and software** (i.e. fixed, wireless, and satellite/hybrid networks) to enable connectivity and distribute secure, ubiquitous and sustainable digital services essential to EU citizens and businesses.
- **Computing and AI**, i.e. infrastructure, platforms and advanced technologies needed to autonomously develop and scale up digital services, enabling companies to innovate,

¹⁰ https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/europes-digital-decade-digital-targets-2030_en

¹¹ https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en

¹² <https://digital-strategy.ec.europa.eu/en/library/staff-working-document-data-spaces>

¹³ <https://digital-strategy.ec.europa.eu/en/policies/cloud-alliance>

¹⁴ https://commission.europa.eu/topics/strengthening-european-competitiveness/eu-competitiveness-looking-ahead_en#paragraph_47059

boost their productivity and upscale, notably concerning cloud, high-performance computing and quantum, as well as AI and its industrial applications.

- **Semiconductors**, a key driver and enabler for the electronics value chain, and a strategic element of Europe's security and industrial strength across sectors.

The European Cognitive Computing Continuum encapsulates this vision and convergence of cloud and IoT technologies, to provide a seamless data processing environment integrated with AI technologies across European cloud service providers. The European market and policy context for the computing continuum are presented in further detail in Section 2A European Perspective: Markets, Laws, and Initiatives. Section 3 presents several trends and European challenges and opportunities for creating the European Cognitive Computing Continuum, which will also be referred to as the *computing continuum*.

This is the second official version of the *European Cognitive Computing Continuum Research & Innovation Roadmap*, and an iteration of the previous draft version of the roadmap from July 2024. The work developing this roadmap was initiated in the winter of 2024 and the project will produce a final version of the roadmap by summer 2026. The process for developing the final version will involve consultation activities to gather feedback from relevant experts and communities.

The analysis of the work, roadmaps, research and strategic agendas, and other relevant outcomes from stakeholders discussed in Appendix A has served to cluster and identify relevant areas of research in the topic of European Cognitive Computing Continuum. To this end, the different initiatives, and the topics of relevance in which these initiatives are working has served as input for the identification of the preliminary subjects to be incorporated into the *Research and Innovation Roadmap*.

1.1 Aims and overview of this document

The aim of the *European Cognitive Computing Continuum Research & Innovation Roadmap* is to identify research and innovation needs and priorities towards achieving this vision, focusing especially on longer term aspects and bridging industry innovation with lower-TRL technologies and excellent academic research.

It is developed as part of the NexusForum.EU project, which is funded by the Horizon Europe programme. The project aims to deliver a strategic outlook on the Cognitive Computing Continuum future, by combining the vision, priorities, and advancements of the IPCEI-CIS and the European Alliance for Industrial Data, Edge and Cloud. On the one hand, the project will produce annual research and innovation roadmaps identifying gaps, opportunities, and international synergies, and on the other it will produce annual research and policy recommendations in key strategic areas.

The wider aim of the NexusForum.EU project is to nurture the European computing constituency, incorporating investors, Member State representatives, and users from different sectors. Finally, NexusForum.EU aims to engage the research and industry ecosystems and relevant initiatives in strategic non-EU countries, with a special focus on Japan and the Republic of South Korea.

This roadmap consists of multiple sections, with recommendations presented in the sections 4 to 9. This roadmap will be further revised and published in a final version in mid 2026, and the recommendations are therefore preliminary and subject to feedback from the community.

Section 2 provides an overview of the European policy and market context for the creation Cognitive Computing Continuum.

Section 3 presents in more detail specific circumstances, challenges, and opportunities that shape the road ahead towards a Cognitive Computing Continuum in Europe.

Sections 4 presents an analysis of transversal aspects regarding the Cognitive Computing Continuum.

Sections 5 to 8 present the identified research and innovation needs clustered as follows, based on early feedback from the European Commission regarding their strategic priorities:

- **AI for Cloud-Edge**, focusing on the challenges when federating compute infrastructure and services from European providers, and for ensuring energy-efficient and secure data processing across the computing continuum.
- **Cloud-Edge for AI**, focusing on the needs and requirements for the computing continuum to enable and support the development and deployment of AI technologies in Europe, based on European values of trustworthy and ethical AI.
- **Telco Cloud-Edge**, focusing on the role of telecom networks and operators in the computing continuum, both as a tenant/user and as a service provider.
- **Digitalisation of Industry Sectors** focusing on the needs, requirements, and challenges for critical sectors and next-generation use cases and applications to use the computing continuum and adopt cloud and edge computing technologies.

Each of these sections, or main topics, are further subdivided into a number of research & innovation subtopics. Each subtopic is presented with a brief introduction, the main challenges, identified research & innovation priorities, and potential impact for Europe and the Cognitive Computing Continuum.

Additionally, section 9 presents several innovative technology areas with potentially disruptive impact on the Computing Continuum in the longer term:

- Neuromorphic systems
- Space edge
- Integration of quantum computing infrastructure
- Quantum and classic computing fusion

Figure 1 below shows all of the subtopics in sections 5 to 9, coloured according to the section it appears in.

In the figure, the y axis indicates a preliminary assessment of the current maturity level of each subtopic, in terms of technology and market adoption. A low maturity indicates a need to further develop the basic technology and market adoption, before it is ready to be adapted to the Computing Continuum. A high maturity indicates that the basic technology is well-developed and has more established market adoption, and that the main challenges are in

adapting it to a multi-provider Computing Continuum. The x axis indicates our preliminary suggestion for when we should expect to start adapting and integrating the technology in the Cognitive Computing Continuum.

The overall interpretation of the figure is that the subtopics that appear closer to the lower left corner, i.e., those with a low maturity level but an early suggested integration timeline, need extra attention to increase the maturity level to meet the expected timeline.

These assessments are preliminary and subject to change in future versions of the roadmap.

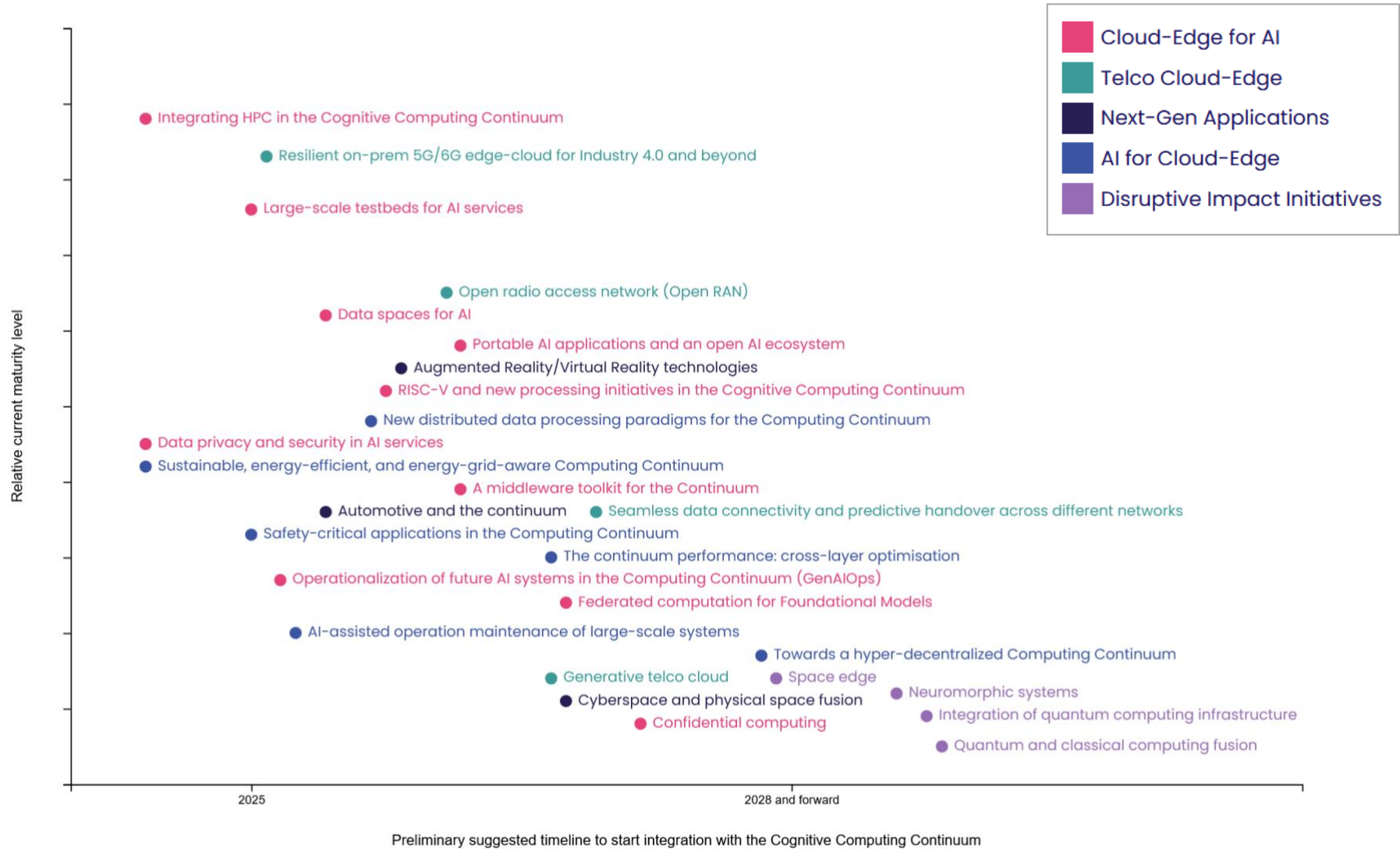


Figure 1 The plot shows, for each subtopic in sections 5 to 9, a preliminary assessment of the maturity level in terms of technology and market adoption, and a preliminary suggested timeline for when to start adapting and integrating the technology in the Cognitive Computing Continuum. The subtopics are colour-coded based on the section it appears in (the main topic it falls under). Subtopics closer to the lower left corner would require extra attention to ensure that the maturity level progresses to meet expectations

2 A European Perspective: Markets, Laws, and Initiatives

Europe consists of multiple countries with diverse regulations, languages, cultural contexts and markets. This fragmentation can hinder the development of unified technological solutions and the emergence of a fully integrated European Digital Single Market. Therefore, harmonizing standards and creating interoperable systems across borders is essential for technological sovereignty. Despite significant initiatives at EU level to harmonize regulatory standards,¹⁵ the digital market in Europe remains fragmented, with local markets individually lacking the critical mass for players to scale and compete with their global counterparts.

The report “*The future of European competitiveness*”¹⁶ assesses the current state of European competitiveness, with a particular focus on industrial policy, its caveats and hopes for its future. Due to geopolitical instability, the urge of not depending on other countries is higher, because the EU has realized that dependency easily engenders instability. According to this report, there has been a shift in paradigm, “The era of rapid world trade growth looks to have passed, with EU companies facing both greater competition from abroad and lower access to overseas markets”. Europe must radically change for digitalisation and decarbonisation to take place in the European economy, investments need to rise to 1960s-70s levels.

The above report further identifies three key areas for growth to take place and to close the innovation gap with the US and China, i) activate a joint plan aimed at strengthening competitiveness ii) decarbonising the economy; iii) take action to enhance security, while reducing dependencies on third parties. In particular, three main barriers prevent Europe from growth.

- Market *fragmentation* is high, which in return drives innovative companies away to more profitable continents. Administrative and regulatory burdens within the European Internal Market prevent innovative companies from thriving.
- The EU is not taking *full advantage* of its common resources. In the field of innovation, collaboration is weak, despite the known spillover benefits of investing in cutting-edge technologies. Here Draghi gives an interesting figure: the public spending on research and innovation in Europe roughly equals the US in terms of GDP share, however, only one tenth of this spending takes place at the EU level.
- Europe does not coordinate on its *industrial strategy*, and its slow and fragmented law-making process hinders its effectiveness in keeping pace with the rest of the evolving world. Overall, the report aims at delivering action points for a new European industrial strategy that can relieve the Union from these barriers.

¹⁵ Relevant examples include the GDPR and the Free Flow of Non-Personal Data Regulation.

¹⁶ https://commission.europa.eu/topics/strengthening-european-competitiveness/eu-competitiveness-looking-ahead_en

The European Union acknowledges the significance of cloud computing, edge computing, and the Internet of Things (IoT) in shaping its digital future toward a European Computing Continuum Infrastructure.^{17,18,19}

The work of NexusForum.EU serves as a facilitation instrument that on one hand assesses the current landscape at the technological, economic and legal levels, and on the other, by engaging with technical experts from both academia and industry, policy makers, and more. In this capacity, NexusForum.EU advises the European Commission on issues related to the European Computing Continuum.

2.1 European market context

Mario Draghi's Report "The future of European competitiveness"²⁰ mentions the lack of innovation as one of the reasons why the EU's industry lacks dynamism. More specifically, it argues that this is due to the lifecycle of this innovation, starting from the initial product idea to its commercialisation. First, companies encounter financial barriers due to the quantity and the quality of funding dedicated to high risk, breakthrough, research and innovation (R&I) areas. In the EU, the capital market is smaller compared to the one of the US, and the venture capital sector is far less developed. Considering the global share of venture capitalist funds, the EU only raises around 5% of these, while the US accounts for more than half of them (52%), followed by China with a 40% share. Therefore, companies willing to scale-up in Europe, often seek alternative growth opportunities abroad, in markets where they are able to have more reach and have higher remunerative returns; in response to this, the demand for venture capital finance in Europe shrinks, alighting a vicious circle. As a matter of fact, Europe lags behind in innovative technologies that have the power to drive productivity growth. With regards to the innovative technology of cloud, the numbers are loud and clear: 3 US tech giants ("hyperscalers") account for more than 65% of the total share of the global cloud market.

If new companies emerge in the EU, apart from funding and venture capital issues, they are also faced with regulatory obstacles that do not accelerate the scale-up phase. Because of this, ideas born in the EU are ultimately more inclined to seek financing from US venture capitalists, also because the US market is less fragmented than the EU one. In Europe, there is not enough public spending devoted to breakthrough innovation. The venture capital invested into breakthrough technologies at the very early development stages is 80% lower compared to US levels, and once the developed technology reaches a later stage, the lack of investment compared to the American counterparts reaches peak gap levels (-82%).

The most prominent funding programme of the EU, Horizon Europe, has an extensive budget but bureaucratic barriers and low focus on breakthrough innovation prevent it from being successful. Indeed, the process of capitalizing on breakthrough technologies is hindered by excessive fragmentation. For example, the EU lacks coordination standards which prevent it

¹⁷ *European Alliance for Industrial Data, Edge and Cloud | Shaping Europe's digital future (europa.eu)*

¹⁸ *European data strategy - European Commission (europa.eu)*

¹⁹ *Europe's digital decade: 2030 targets | European Commission (europa.eu)*

²⁰ https://commission.europa.eu/topics/strengthening-european-competitiveness/eu-competitiveness-looking-ahead_en

from having a real presence in the edge computing field. More start-ups need to scale-up, to enhance the innovation of the European digital landscape. The report suggests that the EU should assist innovative companies by reducing fragmentation. In practical terms, this would mean adopting a new EU-wide legal statute - under the label of “Innovative European Company” - that would equip companies with a single digital entity recognised at supranational level, valid in all the European Member States. In this way, the establishment of subsidiaries across the EU would be facilitated. Companies would be informed about harmonized legislation, especially in the areas of corporate law, labour law and taxation, rendering the whole legal process a positive push rather than an obstacle.

The European Innovation Council Pathfinder²¹ is a European initiative that was meant to spur innovation, although it has a lower budget compared to similar initiatives in the US. Furthermore, it has been criticised due to a lack of scientific experts on the board. Draghi cites, see Figure 2, in 2021, where the share of GDP spent on Research and Innovation was twice for the US compared to the EU. This gap in innovation is directly linked to the gap in productive investment between the European and the US economy that is shown in Figure 2. This is again due to lower investments in technological aspects.

Productive investment

Real gross fixed capital formation excluding residential investment, % of GDP



Source: EIB, 2024.

Figure 2 Productive investment over two decades, EIB 2024

Diving deeper into the issue of the productivity gap that exists between the EU and the US, Draghi identifies the tech sector as a key indicator of why this gap exists. Indeed, around the mid-1990s, Europe did not capitalize on the digital age enough. It did not generate equal amounts of innovative tech companies, nor did it let digital technology empower its economy. The result is an ecosystem made of scarce industrial dynamism and low levels of innovation, investments and productivity.

Europe has a strong foundation to build upon, but the technical roadmap can only be fulfilled in a receptive EU marketplace and research environment. Therefore, the challenges and opportunities outlined in this roadmap need to be co-addressed across the European economy.

²¹ EIC Pathfinder - European Commission

2.2 European legal context

From a European legal context, the digital ecosystem is shaped by several key regulations and initiatives aimed at ensuring a fair, competitive, and secure digital environment. Thus, during the last decade Europe has been developing a regulatory framework aimed at establishing the future European Digital Single Market.²² Ensuring fair competition and fostering innovation, creating a safer digital space, giving individuals greater control over their data, and promoting digital sovereignty, inclusion, equality, and sustainability are among the primary objectives Europe aims to achieve with this regulatory framework.

Within this context, the European Computing Continuum and its associated activities are playing a crucial role in establishing relevant requirements and guiding the needs for the regulatory landscape. Legislation shapes the European Computing Continuum by:

- Ensuring robust data protection measures across all stages of data processing.
- Encouraging the development and deployment of AI systems that are transparent, fair, and accountable.
- Creating a more open and competitive data economy, enabling innovation and collaboration.

These laws collectively aim to create a secure, fair, and innovative digital environment in Europe, balancing the need for data protection with the benefits of data-driven innovation.

The current regulatory landscape in the European Cloud-Edge Continuum is summarised in Figure 3, which shows the timeline and impact of key regulations. These regulations are further described under the figure and can be roughly grouped into regulations aiming to i) create a digital single market in Europe, ii) regulate how data and AI are used in Europe, and iii) safeguard the European digital environment.

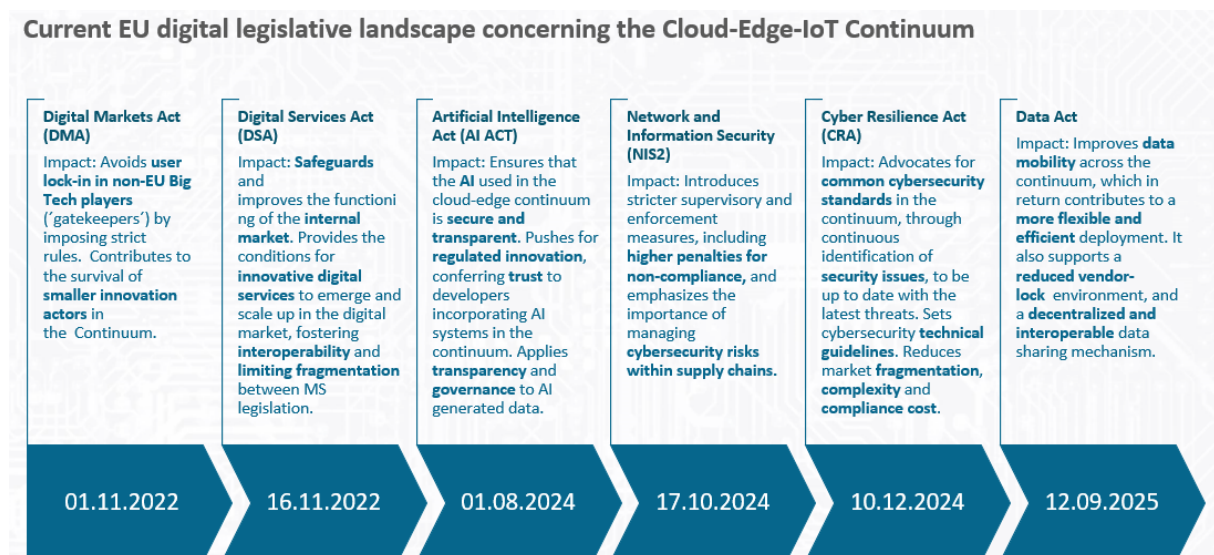


Figure 3. EU laws over time and their impact in the European Computing Continuum.

²² <https://www.consilium.europa.eu/en/policies/digital-single-market/>

The Digital Markets Act (DMA)²³ and the Digital Services Act (DSA)²⁴ are two pivotal regulations in the European Union aimed at creating a fairer and safer digital environment, reducing the dominance of large platforms and promoting a competitive market, enhancing protections for users, fostering a safer online experience and encouraging the growth of smaller companies and startups by creating a fairer market landscape.

A key focus for the European Union is the establishment of a comprehensive legal framework to regulate data in the digital context, ensuring the protection of personal data, promoting fair access, and fostering trust. The General Data Protection Regulation (GDPR),²⁵ the AI Act,²⁶ and the Data Act²⁷ are significant pieces of legislation in the European Union that regulate how data is shared, stored, and processed, and how artificial intelligence (AI) is used within the EU.

Another key objective of the current regulatory framework is to establish a robust system to safeguard the European digital environment. To achieve this, the primary regulations include the Cyber Resilience Act (CRA)²⁸ and the Network and Information Security Directive (NIS2)²⁹ aimed at enhancing cybersecurity across the digital landscape. Together, the CRA and NIS2 contribute to a more secure and resilient digital environment by raising cybersecurity standards, building trust among users ensuring the provision of clear information about the security of digital products and services, strengthening incident response capabilities promoting the detection, reporting, and response to cybersecurity incidents across the EU, and encouraging cooperation between Member States and various stakeholders to address cybersecurity challenges collectively. These regulations are crucial for protecting the integrity, security, and resilience of the EU's digital infrastructure and services.

A key policy document when looking more closely at the IoT sector is the Chips Act.³⁰ The Regulation aims both at monitoring the European semiconductor sector and at supporting the development of *next-generation semiconductor* and *quantum technologies*, thus having a potentially far-reaching impact on the wider IoT technology ecosystem.

In the future, new legislation will also be proposed to address Europe's future challenges in the computing continuum. In her confirmation hearing, Henna Virkkunen, the Commissioner for Tech Sovereignty, Security and Democracy, stated her intentions to progress an EU Cloud and AI Development Act.³¹ The Act aims to "*bring this much needed independent computing capacity for our small businesses*", bridging the gap in data centre premises in the USA and in the EU.

²³ <https://eur-lex.europa.eu/legal-content/EN/TXT/?toc=OJ%3A%3A2022%3A265%3ATOC&uri=uriserv%3A%3A%3A2022.265.01.0001.01.ENG>

²⁴ https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en

²⁵ <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016R0679>

²⁶ <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689>

²⁷ <https://digital-strategy.ec.europa.eu/en/policies/data-act>

²⁸ <https://digital-strategy.ec.europa.eu/en/policies/cyber-resilience-act>

²⁹ <https://digital-strategy.ec.europa.eu/en/policies/nis2-directive>

³⁰ Regulation (EU) 2023/1781

³¹ [virkkunen_verbatimreportheating-original.pdf](#)

The broader aim of the Act would be to reduce technological dependency on foreign providers. The legislation aims to address the EU's growing productivity gap with China and the USA³² by boosting investment in cloud and AI technologies. The spillover effects would be felt by SMEs but also by bigger market players, boosting investments in the subsectors of Cloud and AI, while building decarbonization compliance incentives, to ensure a speedy innovation and green targets can go hand in hand, thus fostering the twin transition objectives as outlined in the Political Guidelines 2024-2029.³³

2.3 European initiatives related to cloud and edge

The European data strategy outlines the steps towards a single European market for data, where data can flow more freely between sectors, and the access and use of data respects European rules. It highlights a need for data processing infrastructure and services that are in line with European laws and regulations.

One key focus in this strategy is the establishment of common European data spaces, meant to pool and share data across sectors. In reverse chronological order, there are four major initiatives that have shaped the data space landscape: SIMPL, GAIA-X, IDS and FIWARE. Each have their takes with SIMPL addressing edge-to-cloud federations and integrating components and solutions from previous initiatives, GAIA-X with a strong focus on transnational platforms/frameworks and compliance checking, IDS providing toolkits and APIs, and FIWARE with its context information management for "smart X" applications.

To address the technical challenges, the EU has launched several initiatives to develop foundational cloud technologies in Europe, support the development of AI in Europe, and incentivise data sharing between organizations in Europe, some of which are:

- The Important Project of Common European Interest on Next Generation Cloud Infrastructure and Services (IPCEI-CIS), recently approved by the European Commission. IPCEI-CIS is a stakeholder in this roadmap.
- The European Alliance for industrial Data, Edge and Cloud³⁴ financed by EU businesses and Member States under the sponsorship of DG CONNECT. The Alliance is similarly a stakeholder in this roadmap.
- Simpl, an open-source smart middleware platform funded by the European Commission, that enables cloud-to-edge federations for the development of Common European Data Spaces.

These are all expanded on in the bullet points below. Also, Figure 4 illustrates how these initiatives fit together with the overall EU vision for cloud and data.

³² <https://www.usnews.com/news/technology/articles/2024-10-23/eus-tech-security-nominee-to-boost-ai-use-battlefield-technology>

³³ https://commission.europa.eu/document/e6cd4328-673c-4e7a-8683-f63ffb2cf648_en

³⁴ <https://digital-strategy.ec.europa.eu/en/policies/cloud-alliance>

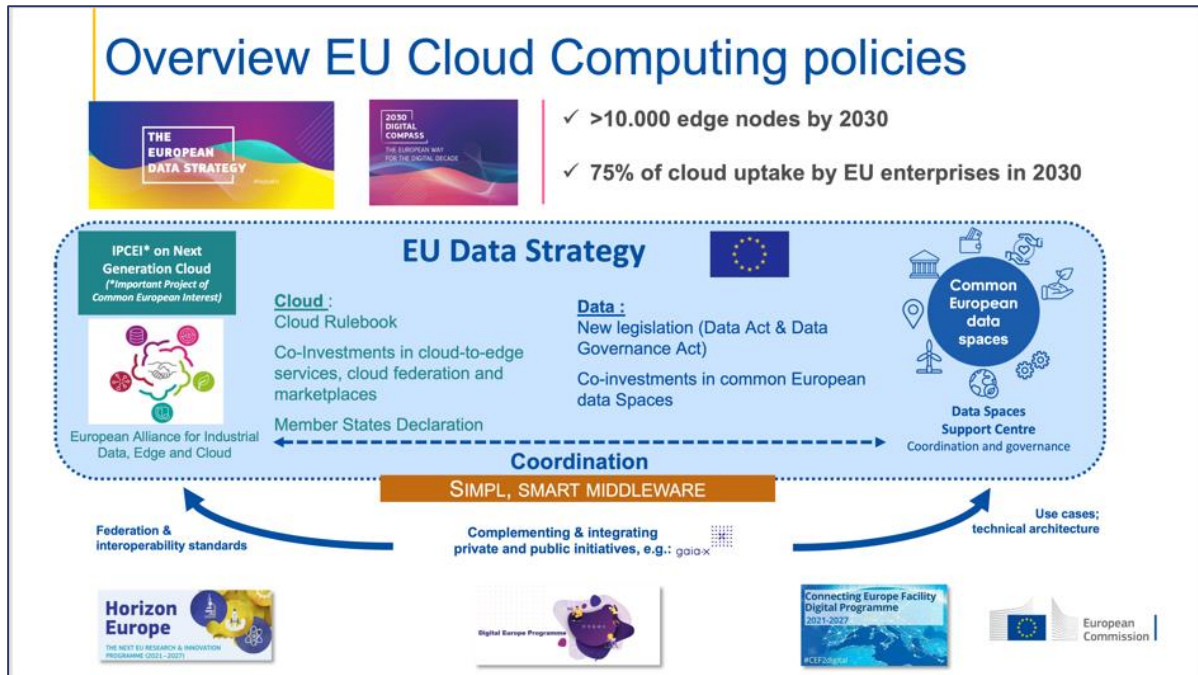


Figure 4 How different European initiatives and policies related to cloud computing fit together. Source: European Commission

Important Project of Common European Interest on Next Generation Cloud Infrastructure and Services (IPCEI-CIS)

At the end of 2023, the European Commission approved the Important Project of Common European Interest on Next Generation Cloud Infrastructure and Services (IPCEI-CIS), the first ever IPCEI focused on cloud technologies; 19 companies from 7 Member States are taking part in it as “Direct Participants”, with a total of around 100 organizations currently active in this initiative.³⁵

The aim of this innovative approach is to foster a multi-provider Cloud-Edge Continuum in Europe, that can create an open environment where data is processed based on a network of interconnected cloud and edge infrastructures.³⁶ This kind of cloud and edge ecosystem is expected, in return, to contribute to the development of data processing technologies that are federated, energy-efficient and trustworthy, thus touching upon European businesses and citizens through the advancement of the EU’s transition towards a green, digital, resilient, secure and sovereign future.³⁷

European Alliance for Industrial Data, Edge and Cloud

EU businesses and Member States have come together under the sponsorship of DG CONNECT to form the *European Alliance for industrial Data, Edge and Cloud*,³⁸ an initiative targeting the development and deployment of next generation edge and cloud technologies,

³⁵ https://competition-policy.ec.europa.eu/state-aid/ipcei/approved-ipceis/cloud_en

³⁶ https://ec.europa.eu/competition/state_aid/cases/1/202412/SA_102517_707E5C8E-0000-C216-8C1C-3081176554C2_287_1.pdf

³⁷ https://ec.europa.eu/commission/presscorner/api/files/document/print/en/ip_23_6246/IP_23_6246_EN.pdf

³⁸ <https://digital-strategy.ec.europa.eu/en/policies/cloud-alliance>

willing to empower the EU industry's position in the field, while fostering effective and secure processing of sensitive data across businesses and public administrations.

The EU Cloud Alliance produces roadmaps to identify priorities for co-investment in cloud and edge technologies, providing recommendations on how to implement such investments and finally, it advises legal and commercial conditions for cloud services. Industry stakeholders willing to join the EU Cloud Alliance must be legally based in the EU and must comply with the overall objectives defined by the EC.³⁹ The work of the EU Cloud Alliance underlines the importance of more technological sovereignty as a crucial approach for the development of a competitive digital European single market.

Simpl

Simpl is the open-source smart middleware platform that enables cloud-to-edge federations and all major data initiatives, in particular the development of Common European Data Spaces in a modular and interoperable way, funded by the European Commission⁴⁰. Common European Data Spaces are a federated data ecosystem based on shared policies and rules. It also works on Data Sovereignty where Data holders remain in control of who can access and use their data, for which purpose and under which conditions. Simpl is made of three parts:

- Simpl-Open: The open-source smart middleware
- Simpl-Labs: The playground environment for Simpl-Open and interoperability test for existing data spaces
- Simpl-Live: instances of Simpl-Open for common European data spaces

The Simpl-Open product will have its first Minimum Viable Prototype (MVP) by the end of December 2024 and will be presented in the first Annual Community Event on January 30, 2025. They are in the phase of continue requirements gathering and their message is *"Reuse if it is available and develop when it is not"*. It is possible to contribute to the community at their website.⁴¹

³⁹ <https://ec.europa.eu/eusurvey/runner/EU-Cloud-Alliance>

⁴⁰ <https://digital-strategy.ec.europa.eu/en/policies/simpl>

⁴¹ <https://simpl-programme.ec.europa.eu/dashboard/community>

3 Towards a European Cognitive Computing Continuum

In order to understand what the Cognitive Computing Continuum, also known as the Cloud-Edge Continuum, is, this introduction will present how the technology could be defined. As well as delimiting the scope, the technical topics can fit in below within such definitions. Let us start with how others see the continuum:

The EU H2020 H-CLOUD project states that:

“The progressive convergence between Cloud Computing and the Internet of Things (IoT) is resulting in a Computing Continuum. The multi-faceted concept of Edge Computing first became to represent a middle ground between data centres and IoT hyper-local networks of sensors and actuators”⁴²

The Federal Ministry for Economic Affairs and Climate Action in Germany:

“The Cloud-Edge Continuum enables greater adaptability to the specific requirements of an application. It offers both the ultra-low latency and real-time processing of edge technology as well as the scalability and resource diversity of the cloud”⁴³

The Compute Continuum can be seen in a multi-dimensional manner. One dimension is from a technical perspective: AI for Cloud-Edge, Cloud-Edge for AI and Telco Cloud-Edge. A second dimension is the Compute Continuum through the lens of Open-Source code, Digital Sovereignty, Interoperability and Cybersecurity, as well as energy consumption and sustainability. A third dimension is markets, rules, policies and regulation. These aspects are particular to Europe, with its stricter rules on the use of personal information, AI and data.

This roadmap outlines some of the missing components for Europe to produce a competitive Compute infrastructure. A “multi-provider computing continuum” represents a seamless integration of cloud and edge computing. Resources and applications are moved both to the cloud and to the edge of the network. This enables dynamic distribution of data processing tasks as required.

There are a number of trends and European circumstances that shape the future European Cognitive Computing Continuum:

- A fragmented European datacentre and cloud market.
- Convergence of networking and compute, and open radio access networks.
- Convergence of operational technologies and information technologies.
- Establishment of a European semiconductor ecosystem based on the RISC-V open standard.
- EuroHPC and the establishment of AI Factories.
- Lack of European data for AI training, and European data regulations.

⁴² <https://h-cloud.eu/>

⁴³ <https://www.bmwk.de/Redaktion/EN/Artikel/Industry/ipcei-cis.html>

These will be explained in more detail below.

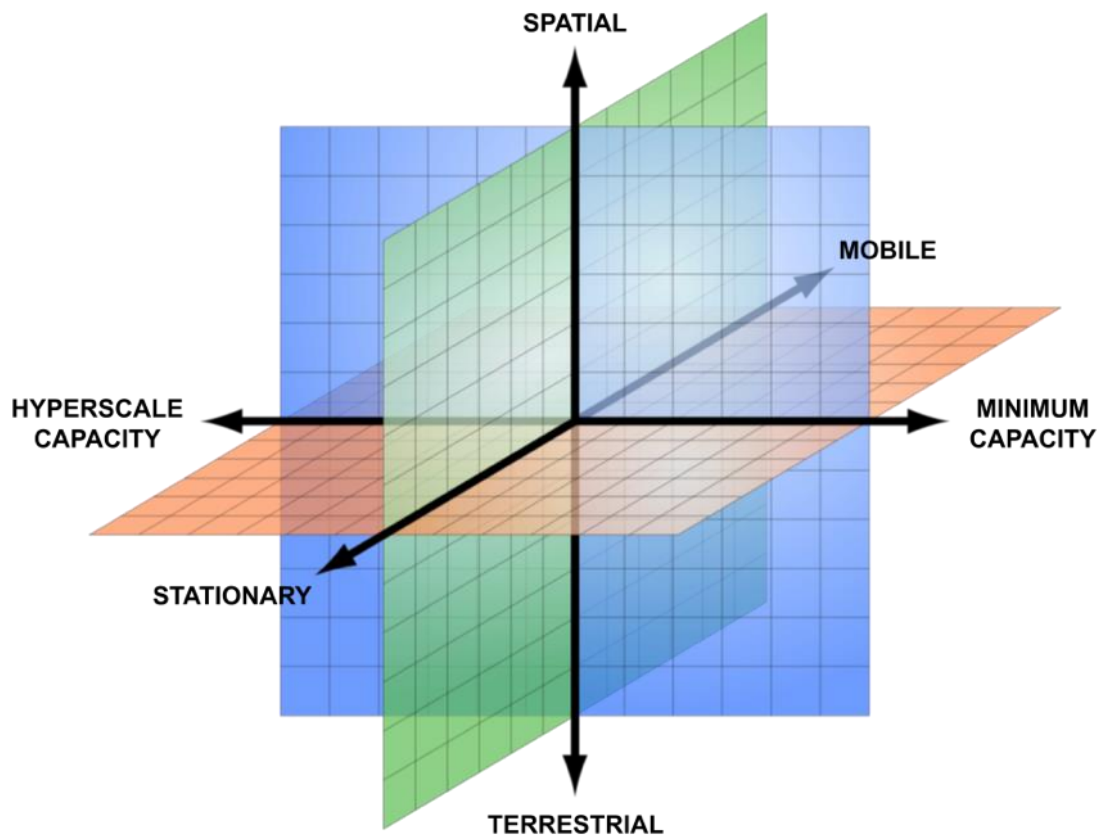


Figure 5 Main dimensions of the emerging Cloud-Edge Computing Continuum infrastructure.

3.1 Fragmented European data centre market

The three largest data centre markets are North America (36%), APAC (31%) and Europe (22%)⁴⁴. The European data centre market's dynamics differ from others, in part due to the prominence of Internet Exchange Points (IXPs). IXPs play a more critical role in Europe, significantly shaping the structure and development of the data centre market. Europe has a dense network of IXPs that facilitate regional and international internet traffic exchange. Major IXPs like DE-CIX (Frankfurt), AMS-IX (Amsterdam), and LINX (London) anchor Europe's connectivity, which influences the location and concentration of data centres in Europe, particularly in IXP-heavy cities.⁴⁵

The European data centre market is fragmented, as evidenced by the diversity of operators and the FLAP-D markets (Frankfurt, London, Amsterdam, Paris, and Dublin) dominate (80% of total power demand in Europe with a recent 17% CAGR and utilisation growing to 88%)⁴⁶, other

⁴⁴ <https://www.precedenceresearch.com/data-center-market>

⁴⁵ <https://www.cbre.com/insights/reports/european-data-centres-overview>

⁴⁶ <https://www.rolandberger.com/en/Insights/Publications/State-of-the-European-data-centre-market.html>

countries like Spain, Italy, and the Nordics are emerging as growth hubs due to favourable regulations, renewable energy resources, and lower operational costs (23% CAGR and utilisation rising to 82%). This geographical diversity demonstrates that no single region monopolizes the market. The fragmented nature of the market allows for competition and innovation but also poses challenges such as inconsistent regulatory environments, varied energy efficiency standards, and differing market maturity levels across regions. This fragmentation requires data centre operators to adopt region-specific strategies and leverage partnerships to thrive. The market is highly competitive without dominant players and 26.68% of the market is represented by five major companies, namely Digital Realty Trust Inc., Equinix Inc., NTT Ltd, SFR and Virtus Data Centres Properties Ltd with other important companies including CyrusOne Inc., Data4, Global Switch Holdings Ltd, Leaseweb Global BV, Stack Infrastructure Inc., Telehouse (KDDI Corporation) and Vantage Data Centres LLC.

The European data centre market is experiencing significant growth, driven by increased demand for digital infrastructure, cloud computing, and data storage solutions. The size of the European data centre colocation market by revenue is estimated to be between \$16.99 billion in 2023⁴⁷ and \$35.46 billion in 2024⁴⁸ with a forecasted CAGR of 13.21% (2023-2028) and 10.2% (2024-2029). One report⁴⁹ forecasts the market (colocation and more) CAGR to be 26.7% (2023-2028) growing to a market size of \$291.7 billion at the end of the period, 2023 -2028. This expansion is fuelled by rising data consumption, the adoption of hybrid and multi-cloud solutions, and the increasing need for high-performance computing, particularly for artificial intelligence and machine learning applications and to certain extent regularity data sovereignty requirements.

The market size by power (IT load capacity) is estimated to be 12.22 GW as of 2024 and estimated to grow to 17.92 GW by 2029 and dominated by tier 3 data centres (with high redundancy level and multiple paths for power and cooling with an uptime of 99.982% and less than 1.6 hours of downtime per year) and this segment is expected to grow due to the adoption of edge and cloud connectivity. The UK market is the largest and hosts the maximum number of tier 3 data centres with Slough and Greater London holding the major share. However, land prices in London are expected to shift investments to other cities such as Amsterdam with its lowest land price in the FLAP-D market. The Spanish market has numerous third-party data centre projects under construction focussed on Madrid and Barcelona. The lack of land availability and skilled workforce in the FLAP-D hubs is expected to help emerging regions in Italy, Poland, Belgium and Sweden. Sweden's market boasts one of the most connected locations in the Nordic region and the country has adopted the waste heat recovery concept for data centres.

⁴⁷ <https://www.globenewswire.com/news-release/2024/02/21/2832512/28124/en/Europe-Data-Center-Market-Overview-and-Forecast-2023-2029-Top-Five-Companies-Hold-25.33-Share-Led-by-CyrusOne-Equinix-Global-Switch-Holdings-NTT-and-Vantage-Data-Centers.html>

⁴⁸ <https://www.mordorintelligence.com/industry-reports/europe-colocation-market-industry>

⁴⁹ <https://www.researchandmarkets.com/report/europe-data-centers-market>

3.2 Fragmented European cloud market

The European cloud market is fragmented, characterised by challenges in scalability, regulatory diversity, and competition from global cloud providers, particularly the US hyperscalers like Amazon Web Services (AWS), Microsoft Azure, and Google Cloud. Europe accounts for 24.7% of the global cloud computing market in terms of revenue in 2023⁵⁰.

However, the investment disparity between Europe and the US in the cloud and ICT sectors, amounting to a significant \$1.36 trillion, presents a formidable challenge for European companies striving to remain competitive⁵¹. To bridge this gap in the next 10 to 20 years, European technology firms would need to dramatically scale annual investments, ranging from \$157 billion to \$1.2 trillion, equivalent to 0.8% to 6.4% of the EU's GDP. Such a leap would demand substantial resource allocation and strategic prioritisation.

Despite commendable initiatives, such as state aid for R&D in cloud and edge technologies, the current funding levels are far below what is necessary to address this shortfall. This underscores the urgency for Europe to adopt open, inclusive policies that attract global investment and foster competitiveness in an increasingly dynamic and technologically advanced global market⁵².

An inward-focused approach to cloud computing and data policies within Europe risks undermining innovation and competitiveness by favouring local champions over global cloud solutions. Such a strategy could reduce market diversity, increase costs, and slow digital transformation across sectors. A more effective approach would involve embracing cross-border cooperation and open market policies. This includes leveraging global expertise while maintaining strong regulatory frameworks to enhance competition, encourage investment, and drive innovation.

The UK's cloud market, though outside the EU post-Brexit, remains an integral part of the European digital ecosystem. It has pursued policies that encourage cloud adoption through initiatives like the G-Cloud framework and a "cloud first" policy in government procurement. These policies highlight the potential for a dynamic regulatory environment that fosters competition and innovation without imposing restrictive localisation requirements. This model could inspire broader European efforts to maintain global competitiveness while addressing local sovereignty concerns⁵³.

Meanwhile, in the EU, initiatives like the European Alliance on Industrial Data, Edge, and Cloud focus on technological sovereignty, aiming to localise data storage and infrastructure. While this approach seeks to retain control over critical assets, it faces significant challenges. For instance, purely European providers like OVHCloud and Deutsche Telekom require massive investments to compete with global giants. Similarly, in the UK, domestic players such as UKFast have struggled to scale in the face of competition from hyperscalers like AWS and Azure.

⁵⁰ <https://www.grandviewresearch.com/horizon/outlook/cloud-computing-market/europe>

⁵¹ <https://ecipe.org/publications/eu-gap-ict-and-cloud-computing/>

⁵² <https://markwideresearch.com/europe-cloud-computing-market/>

⁵³ <https://www.grandviewresearch.com/horizon/outlook/cloud-computing-market/uk>

A successful European-wide strategy must emphasize collaboration, enabling cross-border cloud integration, scaling R&D investments, and fostering innovation. Policymakers across the continent, including the UK, should shift focus from catching up with global leaders to pioneering new technologies and services. This would require policies that encourage interoperability, support competitive markets, and prioritise digital transformation over rigid market share metrics.

Efforts by EU policymakers to establish a self-sufficient ICT infrastructure, exemplified by initiatives like the European Alliance on Industrial Data, Edge, and Cloud, emphasize technological sovereignty. These policies aim to maintain control over critical infrastructure and data while focusing on localisation for emerging technologies like AI and quantum computing. However, achieving this goal would require enormous resources that could detract from other sectors, as purely European firms face substantial challenges in competing with global leaders. A strong indication of this challenge is exemplified by the fact that European cloud providers have grown their revenues by 167% from 2017 to 2022, but their market share of the European cloud market has declined from 27% to 13% as their growth rate lags the market growth.⁵⁴

It is important to note there are private efforts to build a European cloud infrastructure. Some examples include OVHCloud⁵⁵, primarily a French undertaking or the LiDLCloud⁵⁶, from the well-known German supermarket chain. Cloud for compute is being rolled out by companies such as Nebius⁵⁷ and Nscale⁵⁸ in the UK.

3.3 Convergence of networking and compute, and open radio access networks

The convergence of networking and compute will be crucial in transforming Europe's telecommunications sector to support modern complex, data-driven applications. This strategic integration is essential for enhancing network performance by reducing latency, optimizing bandwidth utilization, and improving overall efficiency. This convergence is vital for ensuring that European infrastructure is robust enough to support an economy and society that are increasingly digital. By processing data closer to the source, these converged networks greatly reduce response times and increase the reliability of digital services, fostering technological resilience and operational efficiency across Europe.

The adoption of open network technologies, including aspects of Open Radio Access Networks (O-RAN), plays a part in this transformation by introducing more flexibility and vendor diversity into the network architecture. This approach allows for a more dynamic allocation of resources, adapting in real-time to varying traffic and service demands, which is crucial for the bandwidth-heavy and latency-sensitive applications prevalent today. The O-RAN Alliance⁵⁹ is

⁵⁴ <https://www.srgresearch.com/articles/european-cloud-providers-continue-to-grow-but-still-lose-market-share>

⁵⁵ <https://www.ovhcloud.com/en/>

⁵⁶ <https://horovits.medium.com/lidl-is-taking-on-aws-the-age-of-eurocloud-b237258e3311>

⁵⁷ <https://nebius.com/>

⁵⁸ <https://www.nscale.com/>

⁵⁹ <https://www.o-ran.org/>

defining the leading architecture that exemplifies this broader movement towards open and seamless integration of compute and networking capabilities in radio access networks, but it is important for European sovereignty to also support other approaches to this open networking landscape.

The goal of a fully converged network and compute infrastructure cannot be reached without first addressing some challenges. Technical issues like interoperability between different network components, complexity in managing distributed architectures, and heightened security risks are significant. However, Europe is well-positioned to approach these challenges through joint regulatory frameworks, strategic partnerships across the industry and society, and continuous innovation in cybersecurity and network management.

To remain competitive and secure, the EU need to focus on strategic investments in digital infrastructure, which includes enhancing connectivity through advanced mobile networks and robust data services. This will not only support the immediate needs of European citizens and businesses but can also position Europe as a leading player in the global digital economy.

3.4 Convergence of Operational Technologies and Information Technologies

Operational Technology (OT), such as PLCs and SCADA systems, are widely used for automation and monitoring applications in many industry sectors. These have traditionally used industry-specific hardware components, programming languages and toolchains, and communication protocols, such as Modbus or CAN bus. Each vendor often has its own implementations, leading to vendor lock-in effects, and difficulties upgrading legacy systems as modern technologies become available.

The trend among many vendors in recent years is to offer cloud platforms, connecting their monitoring and automation systems to offer data-driven cloud services and applications. Many have even developed platforms to allow third party vendors to create services and applications that can be integrated with their cloud platforms.

This trend has seen an increased adoption of IoT technologies and modern communication protocols. In addition to the traditional vendors, many SMEs have entered the market to compete with them, providing gateways to integrate legacy systems, offering IoT solutions that include hardware and analytics platforms in the cloud, and other cloud services that can be deployed on top of existing data infrastructure/platforms.

In the future, the expectation is to see further convergence between OT and IT, as adoption of commodity hardware in OT increases, and monitoring and automation systems are further integrated with cloud platforms. Hence, traditional automation systems incorporating modern IoT and cloud technologies could be seen as an emergence of a new edge paradigm.

While the trend of certain OT vendors has moved towards cloud, there could be a very strong use case for the development of an on-prem OT edge paradigm that separates functions between edge and cloud, based on operational constraints. For example, many critical system functions must be maintained even if connectivity to the cloud is lost. Moreover, there are security concerns in use cases with highly distributed automated systems, which is the case for example for building owners that own many buildings spread out geographically.

The convergence of OT and IT technologies should therefore not take the route of moving everything to a central cloud provider but rather seek to combine the strengths of local processing with the peak power of cloud to perform aggregated analytics and advanced AI tasks.

3.5 Establishing a European ecosystem based on RISC-V

Over the past decade and a half there has been significant investment in implementing European processors, developing and maintaining their roadmaps, targeting both HPC and embedded and IoT applications. The EU goal is the production of cutting-edge and sustainable semiconductors in Europe, with at least 20% of world production in value by 2030. By that time, Europe should have manufacturing capacities below 5nm nodes, aiming at 2nm, and improve energy efficiency in the domain by a factor ten. To achieve this the RISC-V open standard Instruction Set Architecture (ISA)⁶⁰ plays a central role in EUs strategy. RISC-V is royalty-free and open-specification allowing any market actor to implement microprocessors or accelerators using it. This lowers the barriers to innovation and market entry, and through collaborative efforts lessens the overhead for individual actors to develop and maintain non-value-adding supporting hardware and software needed for intellectual properties (IP) being developed and brought to market. For this reason, RISC-V has been called “the Linux of hardware”⁶¹. For Europe RISC-V offers the opportunity to develop fully European processors, in contrast to today’s reliance on processors and IPs owned by and developed on other continents.

EuroHPC JU (see section 3.6) aims to build a diverse IP portfolio of processors, accelerators, quantum chips, and AI accelerators, the pilots and demonstrators needed to realize them, as well as the necessary ecosystem to scale them to exascale including the associated software stack and applications. In the short term (2024-2026) this builds on the efforts in the European Processor Initiative (EPI) to build ARM-based processors, but in the medium term (2026-2028) this shall be complemented with competitive RISC-V general purpose processors and GPUs, with EuroHPC JU post-exascale systems being one of the first customers to drive demand in Europe and fulfil the EU goal of autonomy in strategic processing technologies⁶².

The European Processor Initiative, EPI⁶³, is a Framework Partnership Agreement under EuroHPC JU (150 MEUR) to develop European supercomputing technologies, including European microprocessor and accelerator technologies, to improve performance and power ratios. EPI develops general purpose processors based on the ARM ISA, and accelerators based on the RISC-V ISA. The EPI project is complemented by EUPILOT⁶⁴ (30 MEUR) aiming to deliver all-European open-source and open-standard based software and hardware HPC

⁶⁰ RISC-V International, “About RISC-V,” <https://riscv.org/about/>

⁶¹ <https://www.eetimes.com/european-union-seeks-chip-sovereignty-using-risc-v/>

⁶² Alexandra Kourfali (EuroHPC JU), EPI Forum 2024, <https://www.european-processor-initiative.eu/dissemination-material/epi-forum-in-barcelona/>

⁶³ <https://www.european-processor-initiative.eu>

⁶⁴ <https://eupilot.eu>

systems, EUPEX⁶⁵ (40 MEUR) to produce a European pilot for exascale computing, and the eProcessor⁶⁶ (8 MEUR) to develop a European out-of-order high-performance RISC-V CPU.

European processor investment has been targeting both ARM and RISC-V ISAs and while the ARM ISA continues to play an important role in the coming years (2024-2029), there has been a shift towards more investment in European RISC-V based technologies, up to 2030 and beyond, with the establishment of the Digital Autonomy with RISC-V in Europe⁶⁷ (DARE, 240 MEUR) Framework Partnership Agreement to develop large-scale high-performance computing ecosystem based on RISC-V. DARE is expected to run between 2025 and 2030 to deploy RISC-V pilot systems with a path towards post-exascale RISC-V development and procurement in the next decade. These projects do not only leverage technology, IP, and lessons learned from previous European efforts but constitute a coherent and strategic approach to co-develop European RISC-V processors, their software stacks, but also the porting of applications of key European value to the new platforms.

The European Partnership Chips Joint Undertaking, Chips JU, was established following the European Chips Act to succeed the Key Digital Technologies JU with the goal to reinforce EU strategic autonomy and address European challenges in electronic components and systems to establish scientific excellence and innovation leadership through the establishment of pilot lines, design platforms, and competence centres.^{68,69} Chips JU and its predecessors has a strong involvement in European RISC-V which with renewed EU strategic significance today is guided by the 2022 Recommendations and Roadmap for European Sovereignty in Open-Source Hardware, Software, and RISC-V Technologies⁷⁰ and the 2023 Roadmap towards a High-Performance Automotive RISC-V Reference Platform⁷¹. In 2024 Chips JU had contracted two projects in this area; TRISTAN⁷² (54 MEUR, 2022-2025) and ISOLDE⁷³ (39 MEUR, 2023-2026). TRISTAN develops a repository of European RISC-V quality building blocks for SoC designs in key European application domains such as automobile, industrial, as well as the necessary Electronic Design Automation (EDA) tools and the full software stack. The follow-up ISOLDE project expands the European RISC-V ecosystem by maturing the CVA6 and NOEL-V European RISC-V superscalar processors for applications including safety- and security-critical systems, system-level hardware components and their software stacks to contribute to the European RISC-V strategy in embedded high-performance computing and industrial IoT.

⁶⁵ <https://eupex.eu>

⁶⁶ <https://eprocessor.eu/>

⁶⁷ https://eurohpc-ju.europa.eu/specific-grant-agreement-sga-development-large-scale-european-initiative-hpc-ecosystem-based-risc-v_en

⁶⁸ <https://www.chips-ju.europa.eu/Our-vision/>

⁶⁹ <https://digital-strategy.ec.europa.eu/en/news/chips-competence-centres-strengthen-semiconductor-expertise-across-europe-about-kick>

⁷⁰ <https://digital-strategy.ec.europa.eu/en/library/recommendations-and-roadmap-european-sovereignty-open-source-hardware-software-and-risc-v>

⁷¹ Jariu Kinaret (Chips JU), https://eurohpc-ju.europa.eu/document/download/7e21bf39-bb19-43b8-88b6-dd51650d348e_en?filename=pdf%20chips%20parrrt%202.pdf

⁷² <https://tristan-project.eu/>

⁷³ <https://www.isolde-project.eu>

Taken together these European efforts, spanning the Compute Continuum, expose a strong European commitment to a joint European ecosystem based on RISC-V.

3.6 EuroHPC and the establishment of AI Factories

The European High Performance Computing Joint Undertaking (EuroHPC JU) was created in 2018 to make Europe a world leader in supercomputing, with a budget of EUR 7 billion for the period 2021-2027. It is a legal and funding entity that allows the EU and participating countries to coordinate their efforts and pool their resources to develop a European supercomputer ecosystem based on European technologies.⁷⁴

EuroHPC aims to boost scientific excellence and industrial strength in Europe and support the digital transformation of its economy. EuroHPC aims to develop and maintain a *“world-leading federated, secure and hyper-connected supercomputing, quantum computing, service and data infrastructure ecosystem,”* and increase the use of HPC in both public and private sector.⁷⁴

The *AI Innovation Package*, proposed in January 2024, proposed the establishment of AI Factories in Europe, leveraging the EuroHPC supercomputer ecosystem to develop trustworthy AI in Europe.⁷⁵ In July 2024 an amendment to the EuroHPC Regulations came into force,⁷⁶ expanding its mission to include the development and operation of *“AI Factories located around EuroHPC supercomputing facilities to support the growth of a highly competitive and innovative AI ecosystem in Europe.”*

This amendment allows EuroHPC to acquire and operate dedicated AI-optimised supercomputers, in particular for training generative AI and general-purpose AI models that require significant computing resources. The purpose of these AI factories is to offer a one-stop shop for startups, the scientific community, and other innovative AI users to develop powerful AI models for a variety of emerging AI applications.

The EuroHPC AI Factories initiative represents a collaborative effort from across 17 European countries, pooling together EU and national resources to create a robust and interconnected network of AI hubs. This initiative aims to position Europe as a leader in AI innovation and development.

The first seven AI factories will be deployed in 2025 across Europe, specifically in Finland, Germany, Greece, Italy, Luxembourg, Spain, and Sweden.⁷⁷ These sites were selected to host the new AI factories, which will include brand new AI-optimised supercomputers in five of these countries. Notably, some of these will leverage cloud technologies to provide cloud-style interfaces and services that are more familiar to the AI community.

These AI Factories will act as dynamic ecosystems that foster innovation by providing comprehensive support, including access to AI-optimised HPC resources, training, and

⁷⁴ https://eurohpc-ju.europa.eu/about/discover-eurohpc-ju_en

⁷⁵ <https://digital-strategy.ec.europa.eu/en/factpages/ai-innovation-package>

⁷⁶ <https://digital-strategy.ec.europa.eu/en/news/setting-ai-factories-now-possible-after-eurohpc-regulation-amendment>

⁷⁷ https://eurohpc-ju.europa.eu/selection-first-seven-ai-factories-drive-europes-leadership-ai-2024-12-10_en

technical expertise. They will serve as hubs for AI startups, SMEs, and researchers, promoting collaboration across Europe and driving advancements in areas like healthcare, energy, and climate.

3.7 Access to European data, common data spaces, and data privacy

There is an abundance of national and international (EU-level) legislation and initiatives meant to stimulate data sharing across the economy. In the EU, the EU Data Act⁷⁸ mandates equipment manufacturers to make the data collected and produced by the equipment available to users and third parties, supporting ecosystems such as the Common European Data Spaces⁷⁹. Other ongoing initiatives like the International Data Spaces⁸⁰ and Gaia-X⁸¹ aim to bridge the gap between policy, legislation and technical implementation of workable solutions. In Singapore the Infocomm Media Development Authority offers a Trusted Data Sharing Framework and regulatory sandbox to allow businesses and public authorities to experiment with secure data sharing solutions and accelerate innovation⁸². Similarly, in Japan the Ministry of Economy, Trade and Industry and the Information Processing Promotion Agency launched an initiative for interoperable cross-border data infrastructures called the “Ouranos Ecosystem”⁸³. The European Commission funds a wide range of projects and initiatives to implement the European data strategy⁸⁴. For example, European Research initiatives such as the European Open Science Cloud⁸⁵ are working on developing a cloud infrastructure for sharing data for scientific and public services purposes.

In February 2020, the European Commission recognized the need to establish the foundations of a “European Strategy for Data”⁸⁶. Its aim is for the EU to take a leading role in empowering society through data, enabling better decision-making in both business and the public sector. This initiative was followed in May by the Recovery and Resilience Facility’s “20% Digital EU Flagship Scale Up”, in October by the members’ declaration for a “European Cloud,” and in March 2021 by the Digital Decade Strategy, which set targets for Edge and Cloud by 2030.

The data space concept was created to facilitate the sharing of data between European entities. Several initiatives have been launched in Europe to this end:

⁷⁸ EU Data Act <https://digital-strategy.ec.europa.eu/en/policies/data-act>

⁷⁹ Common European Data Spaces <https://digital-strategy.ec.europa.eu/en/policies/data-spaces>

⁸⁰ International Data Spaces Association: <https://internationaldataspaces.org/>

⁸¹ Gaia-X Federated Data Sharing Infrastructure: <https://gaia-x.eu/>

⁸² Singapore Infocomm Media Development Authority: <https://www.imda.gov.sg/how-we-can-help/data-innovation>

⁸³ Japan’s Initiatives for Interoperable Data Infrastructures Officially Named “Ouranos Ecosystem” https://www.meti.go.jp/english/press/2023/0429_001.html

⁸⁴ A European strategy for data <https://digital-strategy.ec.europa.eu/en/policies/strategy-data>

⁸⁵ European Open Science Cloud: https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science/european-open-science-cloud-eosc_en

⁸⁶ <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0066>

- The Data Spaces Support Centre (DSSC), funded by European Commission as part of the Digital Europe Program, establishes a network of relevant organizations and initiatives involved in the development of Data Spaces.
- Gaia-X was established in 2021 as a privately funded not-for-profit organization with the support of BMWK. GAIA-X has funded 11 consortia until 2024, with a total budget of 122 million Euros. While initially focused on building the European Cloud, Gaia-X later opened participation to Cloud Hyperscalers, ensuring their compliance with EU values through a Label Framework.
- The International Data Spaces Association (IDSA), one of the first non-profit organizations to contribute to the European strategy for data, formed in 2017 and incorporated under German law. The IDSA evolved from the Fraunhofer Society project called the Industrial Data Space in 2014, which was renamed to International Data Spaces (IDS) in 2015. IDSA is working on the “Dataspace Protocol,” focusing on the Data Exchange layer of Data Spaces, with the aim to make it an ISO standard. Expanding beyond Europe, the IDSA now has a network of hubs globally, including in Japan, Malaysia, and a competence centre in China.
- FIWARE, originally funded under the European Union’s Seventh Framework Programme for Research, was a non-profit association of industry, academia and SMEs worldwide providing standards and technology development around Data Exchange and Data Sharing and a full framework of open-source components to build smart solutions, digital twins and data spaces.

4 Transversal topics

4.1 Digital sovereignty

Europe's heavy reliance on technologies developed by non-EU vendors, especially those from the United States and China, poses challenges to its ability to control critical infrastructure and data processing services. Meanwhile, non-EU hyperscalers are rapidly advancing in the race to dominate the future computing value chain. Although European cloud offerings span a wide spectrum of services, customers often need to collaborate with multiple providers to match the quality and breadth of services offered by leading global cloud providers. European providers differentiate themselves by focusing on specific niches where they excel. European cloud providers (including telcos) are in a unique position to be able to leverage proximity, resource optimisation, QoS evaluation, and regulatory compliance to enhance edge computing performance and address latency challenges.

Securing Europe's strategic autonomy in the digital world is not only a technical hurdle; it also presents a geopolitical puzzle. Europe contends with other global regions in the race for becoming technological leaders. Navigating the delicate balance between economic interests, security imperatives, and strategic partnerships demands thoughtful management and governance. The impact of this process goes beyond individual choices by end-users and businesses, as illustrated by Ursula von der Leyen's February 2020 op-ed⁸⁷, where she described tech sovereignty as "the capability that Europe must have to make its own choices, based on its own values, respecting its own rules."

For instance, ensuring data privacy and security is crucial for maintaining sovereignty. European citizens and businesses rightfully anticipate that their data will be treated with care and shielded from unauthorized access. Balancing the imperative of data sharing for innovation with the need to protect privacy remains a formidable challenge. While regulations like the General Data Protection Regulation (GDPR) aim to tackle this issue, their implementation across the EU Member States can be intricate and geopolitical threats are also an issue to be addressed.

Furthermore, in 2022, ENISA was commissioned to include sovereignty requirements into the EUCS. These requirements requested that CSPs of "high" assurance level have their headquarters in the EU, and "not be controlled by any non-EU entity" and be "completely independent from non-EU laws". After almost two years of controversy, the latest news⁸⁸ indicate that cybersecurity related sovereignty requirements are likely to be removed from the final EUCS proposal with the aim to ensure fair access to EU market for non-EU CSPs. Whether this is a profitable decision for European Cloud market is still to be seen, and in any case it most likely will significantly impact the Cloud market evolution in the EU. Mechanisms to monitor the business benefits of CSPs whose headquarters reside within EU vs. abroad should be established.

⁸⁷ https://ec.europa.eu/commission/presscorner/detail/es/ac_20_260

⁸⁸ <https://www.euractiv.com/section/data-privacy/news/sovereignty-requirements-for-cloud-providers-will-not-make-it-to-commissions-proposal-for-implementing-act/>

4.2 Open Source

Open technologies, including standards, open-source software (OSS) and hardware (OSH), contribute to economic growth and innovation, and they hold significant potential in strengthening European digital sovereignty.⁸⁹ Recognising that foundational Cloud-Edge technologies are currently, to a large extent, being developed and maintained within opensource foundations like the Linux Foundation and the Eclipse Foundation, the roadmap of the *European Alliance for Industrial Data, Edge and Cloud* outlines several priorities aimed at improving European capabilities and autonomy in the global open-source sector, including the need to "strengthen existing (and create new) open source software communities, led by European organisations, to manage and maintain important technologies in the long term as well as to lead open source reference implementations".⁹⁰

While the critical importance of OSS for cloud computing is well established, interest in OSH has developed more recently and is centred on the role that it can play in lowering barriers to entry and thereby mitigating strategic dependencies in the fields of processors and, more broadly, semiconductors. Such dependencies were made salient by disrupted supply chains during the COVID-19 and increased geopolitical volatility in recent years. Currently, the growing role of AI across various sectors is further heightening the strategic importance of semiconductors and the need for the EU to address particular areas of strategic weakness, including chip design.⁹¹

Of particular interest in this context is the surge in open innovation, catalysed by the development and adoption of RISC-V, a free and open instruction set architecture (ISA) standard. Initiatives like the Chips for Europe initiative, supported by legislations such as The European Chips Act,⁹² is providing an impetus for increased collaboration within Europe, emphasising the importance of RISC-V and open-source. Building on open and shared standards, open-source licenses facilitate seamless collaboration between researchers and industry, helping to accelerate innovation and counteract dependencies. However, to fully harness the potential of open technologies in achieving digital sovereignty, a strategic approach is required to build and strengthen competencies in key areas, ensuring Europe has the capacity to fork projects and maintain the code it relies on, thereby reducing dependencies.⁹³

⁸⁹ Blind et al. (2021). *The impact of Open-Source Software and Hardware on technological independence, competitiveness and innovation in the EU economy. Final Study Report.* <https://digital-strategy.ec.europa.eu/en/library/study-about-impact-open-source-software-and-hardware-technological-independence-competitiveness-and>

⁹⁰ <https://digital-strategy.ec.europa.eu/en/news/european-alliance-industrial-data-edge-and-cloud-presents-its-first-deliverables>.

⁹¹ Kleinhans (2021). *The lack of semiconductor manufacturing in Europe. Policy Brief.* SNV Berlin. <https://www.stiftung-nv.de/de/publikation/lack-semiconductor-manufacturing-europe>

⁹² EC (2022). *European Chips Act: Communication, Regulation, Joint Undertaking and Recommendation.* *European Chips Act: Communication, Regulation, Joint Undertaking and Recommendation*, <https://digital-strategy.ec.europa.eu/en/library/european-chips-act-communication-regulation-joint-undertaking-and-recommendation>

⁹³ EC Working Group on OSH and OSS (2022). *Recommendations and roadmap for European sovereignty on open-source hardware, software and RISC-V Technologies*, <https://digital->

As an example, the report *“Open Source Software in Automotive and its transition towards Software Defined Vehicles”*⁹⁴ presents a vision for open source in the automotive industry. While the conclusions target mainly applications in automotive, they are far-reaching and apply more generally:

- *Better standardisation through EU organisations:* Leverage open source, as in IT as de-facto solutions. Ensure that any standards promoted are implementable in open-source software (OSS). Implement educational initiatives about the benefits and best practices of OSS for businesses, particularly in sectors like automotive.
- *Establish functional safety norms:* Collaborate with industry consortia and stakeholders to develop clear guidelines and standards for functional safety certification that consider and leverage the benefits of OSS characteristics. Cultivate environments that promote collaboration among OEMs, suppliers, and OSS communities to drive innovation.
- *Consider OSS as a geopolitical concern:* Take proactive steps to understand the geopolitical landscape of software and its implications for the industry. Adopt and support open technologies, including OSS, open standards, and open hardware, both in industry and government, to enhance regional sovereignty and global competitiveness.
- *Engage with global OSS sustainability efforts:* Governments should engage in and encourage industry participation in global OSS sustainability efforts to ensure the longevity and support of OSS. Support may include diverse options such as funding critical projects, promoting OSS in public procurement, raising awareness and knowledge of the risks related to the global supply chain, and encouraging organisations to sustain their own upstream dependencies.

4.3 Sustainability and Energy Efficiency

The European Union is increasingly prioritising sustainability and energy efficiency as central components of its digital transformation. This emphasis is enabled by Europe’s ambitious climate goals, regulatory frameworks, and the growing demand for environmentally responsible computing solutions.

The EU has positioned itself as a global leader in sustainable digital transformation with policies such as the European Green Deal and the Climate Law, and strong intentions for climate neutrality by 2050. The computing ecosystem, including data centres and cloud providers, are expected to play an important role in achieving these targets.

The EU Code of Conduct (CoC) for Energy Efficiency in Data Centres⁹⁵ provides voluntary guidelines for reducing energy use. It promotes best practices such as optimising cooling

strategy.ec.europa.eu/en/library/recommendations-and-roadmap-european-sovereignty-open-source-hardware-software-and-risc-v

⁹⁴ Found on LinkedIn; (27) *Open Source Software in Automotive and its transition towards Software Defined Vehicles* | LinkedIn and also at *Open Source Software in the Automotive Industry*

⁹⁵ <https://e3p.jrc.ec.europa.eu/communities/data-centres-code-conduct>

systems, increasing server utilisation, and integrating renewable energy sources as well as aspects of waste heat recovery.

The recast Energy Efficiency Directive (EED) establishes critical requirements for data centres within the EU⁹⁶ and operates as part of a broader effort to enhance sustainability and energy security. By aiming for the 2030 energy efficiency target, the recast EED seeks to reduce energy costs and achieve substantial reductions in greenhouse gas emissions. These objectives are closely tied to the adoption of transformative technologies, including numerous innovations essential for the green transition. The EU CoC for Energy Efficiency in data centres will become part of the framework to audit data centres for the deployment of energy efficient practices.

Improved energy efficiency also plays a pivotal role in strengthening the EU's energy supply security. By decreasing reliance on fossil fuel imports from third countries, the Directive addresses vulnerabilities that have been highlighted by recent geopolitical tensions.

This dual focus on sustainability and resilience ensures that the EU's energy systems are not only greener but also more robust in the face of growing external challenges. The strategic alignment of the EED with other policies under the European Green Deal reinforces its role in achieving climate neutrality by 2050 while ensuring that data centres contribute to this overarching goal through efficiency-driven operations.

Data centres are integral to Europe's computing ecosystem but are also significant energy consumers. As of 2023, they account for roughly 2-3% of the EU's total electricity demand, with projections to grow due to expanding cloud services, AI applications, and edge computing.

To combat this, the EU has introduced measures to increase energy efficiency:

- By pushing for renewable energy integration. Many data centres in Europe now operate on renewable energy. For instance, Equinix and OVHcloud have pledged to use 100% renewable energy across their European operations, with OVHcloud claiming 92% Renewable Energy Factor in 2024⁹⁷ and Equinix claiming 96% use of renewables in 2024.⁹⁸
- By incentivising and mandating waste heat recovery, Nordic countries, such as Sweden, Finland, and Denmark, lead in recycling waste heat from data centres into district heating networks, showcasing an innovative approach to energy efficiency. The EED recast mandates the data centres with greater than 1MW of IT load should deploy waste heat recovery with a specific Energy Reuse Factor (ERF) and only not to enforce this if it is demonstrated to be technically or financially infeasible.

The European digital infrastructure ecosystem is embracing innovative technologies to reduce environmental impact, such as:

- Liquid Cooling Systems. These systems are more energy-efficient from the cooling perspective than traditional air-cooled systems and are increasingly adopted in HPC (high-performance computing) and AI data centres.

⁹⁶ https://eur-lex.europa.eu/eli/reg_del/2024/1364/oj

⁹⁷ <https://corporate.ovhcloud.com/en-gb/sustainability/environment/>

⁹⁸ <https://www.equinix.co.uk/data-centers/data-center-excellence>

- AI for Energy Optimisation, where AI and machine learning are used to predict and manage energy use, thus enhancing efficiency. Google's data centres in Europe, for example, supposedly leverage AI to optimise their cooling systems.
- Edge Computing that enables the processing of data closer to the source, edge computing reduces the need for long-distance data transfers, cutting energy costs and improving latency for certain applications that require very fast response.

Cloud computing, which constitutes a significant part of the European digital ecosystem, offers inherent sustainability benefits compared to on-premises solutions. These include:

- Resource Efficiency, coming from a shared infrastructure in cloud environments, which therefore enables better utilisation rates and lower energy per unit of computation compared to traditional setups.
- Commitments from Providers, for example leading cloud providers like AWS, Microsoft Azure, and Google Cloud have data centres in Europe that are claiming to be carbon-neutral or powered by renewable energy. Local players such as OVHcloud and Atos/Eviden are also investing heavily in green energy solutions.

Despite the progress, the European digital ecosystem faces several significant challenges, namely:

- Balancing Growth and Sustainability with the exponential growth in data usage, driven by AI, IoT, and 5G, risks outpacing sustainability efforts, especially against an electrical grid that is becoming more constrained in places.
- Standardisation and regulation. Even though voluntary initiatives like the EU CoC exist, there is a need for standardised, enforceable sustainability regulations across Member States, although this is currently being attempted for data centres by both the recast EED and the CSRD reporting requirements.
- E-waste Management is also becoming a strong requirement. In this case, upgrading data centres often generates electronic waste and therefore good strategies for sustainable disposal and recycling currently remain underdeveloped.

Opportunities include expanding green energy partnerships, adopting circular economy principles in hardware production, and fostering cross-border collaboration to share best practices and technologies.

4.4 Cybersecurity

One of the major milestones in the cybersecurity landscape in Cloud will be the adoption of the European Union Cybersecurity Certification Scheme for Cloud Services (EUCS)⁹⁹ when it is finally ready and approved. The EUCS is one of several schemes framed within the EU cybersecurity certification framework, as set out in Regulation (EU) 2019/881¹⁰⁰ While adopting EUCS is voluntary, the NIS 2 Directive, in force since October 15, 2024, allows EU MS mandating

⁹⁹ <https://www.enisa.europa.eu/publications/eucs-cloud-service-scheme>

¹⁰⁰ https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:JOL_2013_165_R_0041_01&qid=1397226946093&from=EN

that essential and important entities use only EUCS-certified cloud services. The scheme defines sets of common reference security requirements for different assurance levels of cloud services, ranging from “basic” to “high+”, through ‘substantial’ and ‘high’.

The latest draft of the EUCS was released in November 2023 and its approval is suffering significant delays. The draft EUCS involved several key updates, most importantly the removal of the data localisation requirement for basic and substantial assurance levels, leaving it to only CSPs seeking for “high+” assurance level the obligation of storing European citizens’ data within the EU and establishing contracts governed by the laws of the EU or a Member State.

At the practical level, security processes and controls in the cloud will need to go strengthening as the threat diversity and virulence intensify. For example, solutions that enable seamless security will raise in popularity, such as Multicloud Key Management as a Service (KMaaS) solutions that enable the handling of secrets (e.g., SSH keys, API keys and certificates) in the cloud, both in hybrid environments and environments where multiple cloud providers are used. Other zero trust technologies will also be crucial in the protection of Clouds, enabling the evolution from add-on and reactive security to security-by-design and multi-layered distributed security of networks, edges, devices and servers.

In the short term, Artificial Intelligence (AI) and Generative AI will make Cloud cybersecurity even more challenging since they will be the means for sophisticated and combined attacks, malicious content generation, raising automation and complexity of multiple types of attacks such as malware, ransomware, APTs, etc. In parallel, these technologies are also expected to boost the intelligence capabilities of preventive security and response automation. Therefore, they will play a dual role, aiding in making both offensive and defensive solutions smarter.

Latest years have witnessed a vertiginous increase in the number of IoT devices connected all over the globe. With Cyber Physical Systems (CPS) and Cyber Physical Systems of Systems (CPSoS) gaining momentum with a myriad of related applications, it becomes fundamental to research on light yet effective SIEM solutions that can be embedded / onboarded in those devices that usually present challenging hardware constraints., Additionally, it is important to make a leap forward by defining and extensively testing efficient SIEM hierarchical multi-tier architectures where SIEMs of growing capacity and complexity are placed as we move from lower, IoT, to upper layers, edge and cloud.

As reported above, cybersecurity is expanding at the same speed as the digital concept does, computing continuum (Cloud, edge, IoT), software, hardware, etc. Therefore, a holistic and systems-centric approach is essential, encompassing not just computing, networks, infrastructure, hardware, software, data, platforms, people, and services. This approach highlights the importance of understanding ‘cascading risks’ and ‘systematic risks’ and how to overcome them ensuring that all potential vulnerabilities are addressed to prevent widespread damage. Cyber capabilities by any and all malicious actors threaten all segments and sectors. In this Digital age access to affordable computing power, data, AI capabilities, and the collaboration among cybercriminals have led to the industrialization of cybercrime, making it one of the fastest-growing markets. However, industries, users, consumers, and the public sector are still not taking cybersecurity seriously enough. A major issue in addressing

these significant challenges is the lack of effective implementation, support to the stakeholders and collaboration.^{101,102,103}

In a longer term, *homomorphic encryption* will see practical implementations in the Cloud, allowing for confidential processing of both sensitive (confidential, secret) and private data without the need of decrypting them. While major limitations are currently being addressed¹⁰⁴, such as speed and costs, homomorphic encryption is a promising technology which will make it possible that applications run on top of encrypted data that can only be decrypted by the owner of the secret key.

As it gets more and more mature, quantum computing technology will also be beneficial for Cloud services as a means to accelerate the cryptographic processing in all those communication and data protection algorithms that require high compute power.

In a quantum-enabled cloud, quantum computing can potentially be used to enhance cloud security protocols and privacy. For example, quantum key distribution (QKD) could be used to create very strong cryptographic keys that can be used to establish secure communication channels theoretically immune to eavesdropping. Hybrid security models that combine the strengths of classical and quantum cryptographic techniques to create secure cloud environments will also be possible.

Since the powerful quantum computers may in the future be able to break many of the currently used encryption algorithms to protect data in the cloud, quantum-resistant technologies and solutions are currently being researched. Therefore, it is expected that post-quantum cryptographic algorithms and protocols will also be increasingly adopted to avoid quantum threats to encryption and ensure secure communications and storage in the cloud. In this line, the initial public draft of the NIST IR 8547 Transition to Post-Quantum Cryptography Standards¹⁰⁵ issued November 12, 2024 has already pioneered the path towards clarifying the rationale and challenges of the change.

Real-time cyber risk monitoring systems combine information about vulnerabilities in the target infrastructure, information about threats and attacks detected by SIEM systems, the business profile of the company and the involved digital assets with the purpose of measuring the economic exposure related to the cyber climate in which such an infrastructure is operating. Research could be conducted in evolving the state of the art of the cyber risk models used to calculate this economic exposure so that they can take advantage of the valuable information generated by different monitoring and detection tools. This information is larger and larger in volume and contains insights that are fundamental in obtaining close-to-reality economic cyber risk estimations. Fine-tuned cyber risk monitoring systems are extremely useful when estimating the price of cyber insurance, in the case when a company decides to transfer the cyber risk to a third party by paying a periodic fee.

¹⁰¹ <https://digital-strategy.ec.europa.eu/en/news/commission-calls-23-member-states-fully-transpose-nis2-directive/>

¹⁰² <https://www.enisa.europa.eu/news/eus-first-ever-report-on-the-state-of-cybersecurity-in-the-union>

¹⁰³ <https://www.consilium.europa.eu/en/press/press-releases/2024/12/02/cybersecurity-package-council-adopts-new-laws-to-strengthen-cybersecurity-capacities-in-the-eu/>

¹⁰⁴ <https://medium.com/@allen.westley/homomorphic-encryption-in-action-f8b7e11b885e>

¹⁰⁵ <https://nvlpubs.nist.gov/nistpubs/ir/2024/NIST.IR.8547.ipd.pdf>

Internet of Behaviour (IoB) is based on the detection of behavioural patterns both in the network and the infrastructure that can potentially be mapped to threats. IoB seems highly promising due to the fact that it is not necessary to access data themselves, just detect an anomalous behaviour in comparison to what would be considered a normal pattern. It allows anticipation in risk management and the chance to put in place mitigation measures before an actual attack takes place.

The potential of IoB is connected to Cyber Threat Intelligence (CTI) and in particular to sharing threat information to third parties, modelling the behaviour and building a common knowledge base that permits the detection of long-term attacks. Since its very conception, CTI has faced the issue of reluctance of organizations to share information labelled as delicate, since it is related to threats and attacks on their digital assets. Working groups like MISP have tried to provide answers to this challenge. It is important to foster research on how to guarantee that information is shared in a fully anonymous way to encourage more organization to participate and strengthen common knowledge.

4.5 Interoperability

In open and fragmented business ecosystems which foster innovation and competition, the systems capability to collaborate, being interchangeable and evolve together is a key success factor. This set of capabilities are part of the interoperability concept, which is a side benefit of long-term and occasional collaboration initiatives between organisations. Interoperability is an open and broad term that covers different collaboration aspects, such as data exchange, communication, process integration, and system compatibility.

The interoperability of digital systems is a cross-cutting concern in the European Union that has been materialised through different initiatives and regulations.

From the *EU policy perspective* several initiatives have become relevant to foster interoperability in the Computing Continuum.

The European Interoperability Framework (EIF)¹⁰⁶ is a set of recommendations for promoting interoperability in the delivery of European public services. The EIF was first published in 2004 and the latest version, EIF 2.0, was published in 2017 and provides a set of guidelines for promoting interoperability in the delivery of European public services.

The European Interoperability Act¹⁰⁷ is a regulation that aims to promote interoperability between, and with, European public administrations by establishing a legal framework for the exchange of data and information between different systems and platforms. It addresses issues in the application of the European Interoperability Framework (EIF) and its limitations due to their voluntary nature. European Entry Exit System (EES)¹⁰⁸ results from the European Union initiatives to promote the interoperability of the members' large-scale IT systems in the field of borders, visa, police and judicial cooperation, asylum and migration.

¹⁰⁶ https://ec.europa.eu/isa2/eif_en/

¹⁰⁷ <https://interoperable-europe.ec.europa.eu/interoperable-europe/interoperable-europe-act>

¹⁰⁸ https://ec.europa.eu/home-affairs/what-we-do/policies/borders-and-visas/smart-borders/entry-exit-system_en

At the domain-specific level, various initiatives have been launched to enhance interoperability within each vertical sector.

In the energy sector, the European Commission launched the Bridge initiative¹⁰⁹ to promote continuous sharing of information among research projects. As a result, a supporting topic 'interoperability community' was included as a call for proposals in the Horizon Europe programme.¹¹⁰

In the *air traffic sector*, the Regulation (EC) No 552/2004 on the interoperability of the European air traffic management network, and aims to promote the single European sky by establishing a legal framework for the exchange of data and information between different systems and platforms.¹¹¹ It's worth mentioning that this regulation was repealed by the Regulation (EU) 2018/1139 on common rules in the field of civil aviation and establishing a European Union Aviation Safety Agency. This new regulation also addresses the interoperability of the European air traffic management network.¹¹²

In the *rail sector*, the European Union, while keeping the high level of safety, is promoting the interoperability of the rail system to facilitate the entry of new operators and the development of new services. This is mainly regulated by the directive 2016/797 of the European Parliament and of the Council of 11 May 2016 on the interoperability of the rail system within the European Union.¹¹³

In the *health sector*, under the EU4HEALTH programme¹¹⁴ the European Commission and the World Health Organization (WHO) announced the 11 of December 2023 that they signed an agreement to launch a four year project to strengthen health information systems and boost health data governance and interoperability in Europe.¹¹⁵ In the **food sector**, the European Food Safety Authority (EFSA) and the German Federal Institute for Risk Assessment (BfR) are developing a Universal Traceability data eXchange (UTX) and a multi-actor tracing software ecosystem to harmonize tracing data across Europe.¹¹⁶

Other initiatives have been originated by non-public bodies such as:

- In the financial sector, the 6 of August of 2024, the European Round Table for Industry (ERT) sent a letter to the European Union Commissioner for Financial Markets, Mairead McGuinness, on the need of interoperability between sustainability reporting frameworks.¹¹⁷

¹⁰⁹ <https://h2020-bridge.eu/>

¹¹⁰ https://cordis.europa.eu/programme/id/HORIZON_HORIZON-CL5-2021-D3-01-03/en

¹¹¹ <https://eur-lex.europa.eu/EN/legal-content/summary/interoperability-of-the-european-air-traffic-management-network.html>

¹¹² <https://www.atc-network.com/atc-news/easa-european-aviation-safety-agency-germany/single-european-sky-easas-air-traffic-management-responsibility-consolidated-through-new-regulations>

¹¹³ <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016L0797>

¹¹⁴ https://ec.europa.eu/health/funding/eu4health_en

¹¹⁵ <https://ec.europa.eu/newsroom/sante/newsletter-archives/49640>

¹¹⁶ <https://link.springer.com/article/10.1007/s00003-024-01522-8>

¹¹⁷ <https://ert.eu/documents/ert-letter-to-commissioner-mairead-mcguinness-on-the-need-for-interoperability-between-esrs-and-issb-standards/>

- Eurocities, which is a network of more than 200 European cities, presented and Interoperability Framework for electric charging related technology.¹¹⁸

In the European computing sector, it is crucial to first understand the regulatory priorities regarding interoperability in Europe. These regulations aim to address the European social and economic needs. In this context, the Cloud Computing Policy¹¹⁹ foresees the rise of cloud and edge computing for data processing. Among other aspects, they anticipate an increase of edge nodes in the coming years until 2030, when 10,000 edge nodes are expected to be deployed. This paradigm shift will require *interoperability-oriented initiatives to reap the benefits and avoid the drawbacks of increased fragmentation*. The foreseen interoperability needs are addressed by some of the measures initiated by the European Commission:

In the European Data Strategy, the objective of the Data Act¹²⁰ is to make it easier to switch between different data service providers and achieve free flow of non-personal data. This will require interoperability among different cloud service providers.

In the European Industrial Strategy, the *European Alliance for Industrial Data, Edge and Cloud*¹²¹ aims to bring together relevant stakeholders from the private and public sector to jointly define strategic investment roadmaps to enable the next generation of highly secure, distributed, interoperable and resource-efficient computing technologies.

Finally, it is worth mentioning that there are other acts such as the Digital Services Act¹²² (DSA) and the Digital Markets Act¹²³ (DMA) that may have applicable interoperability requirements. For example, the DMA establishes an “Obligation for gatekeepers on interoperability of number-independent interpersonal communications services”.¹²⁴

¹¹⁸ <https://eurocities.eu/resources/interoperability-framework-for-electric-charging-related-technology/>

¹¹⁹ <https://digital-strategy.ec.europa.eu/en/policies/cloud-computing>

¹²⁰ <https://digital-strategy.ec.europa.eu/en/policies/data-act>

¹²¹ <https://digital-strategy.ec.europa.eu/en/policies/cloud-alliance>

¹²² https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en

¹²³ https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-markets-act-ensuring-fair-and-open-digital-markets_en

¹²⁴ https://www.eu-digital-markets-act.com/Digital_Markets_Act_Article_7.html

5 AI for Cloud-Edge: Orchestration and Managing a Multi-Provider Continuum

Creating a federated, energy-efficient, and trustworthy European data processing ecosystem, spanning a multi-provider cloud to edge continuum, requires the development of new automated management technologies for the Cloud-Edge stack and new mechanisms for interoperation between different providers.

The management of the Cognitive Computing Continuum should enable real-time and low-latency services by distributing workloads in the Computing Continuum and reduce transmission of large volumes of data to centralised cloud data centres. The Computing Continuum needs to be energy-efficient and sustainable, with as low carbon footprint as possible. Data-driven methods, machine learning, and AI should be further developed and deployed to learn patterns and automate the management of the Compute Continuum.

Six topics are identified below, focusing on federating compute resources in Europe and on the foundational technologies needed to make it happen and importantly make the Computing Continuum operate efficiently.

5.1 Sustainable, energy-efficient, and energy-grid-aware Computing Continuum

The rapidly growing adoption of cloud-edge computing and AI services across Europe will lead to a sharp increase in energy demand of data centres in the coming years. As the global share of energy used by data centres increases, the carbon footprint of the computing continuum will play an increasingly important role in achieving the ambitious climate goals, see section 4.3 “Sustainability and Energy Efficiency”.

The management and operation of the computing continuum needs to become not only more energy-efficient, but also carbon-aware. The electricity consumption of data centres in the EU accounted for almost 4% total electricity demand within EU in 2022, at an estimated total of almost 100 TWh. Electricity consumption in the data centre sector in the EU is expected to reach almost 150 TWh by 2026.¹²⁵ This is at a time when decreasing energy use and emissions have been mandated by the EU and most national governments. The energy consumption by the edge, although not trivial to calculate, is estimated at some 2% of that by data centres in 2018, but expected to grow to 12% by 2025.¹²⁶

The role of a holistic data centre management system, overseeing both hardware and software components, ensures that every aspect of the data centres’ operation is aligned with efficiency goals. Traditionally, data centres have operated under the paradigm of maintaining the lowest possible temperature to safeguard hardware integrity. However, with real-time insights provided by digital twins into the actual cooling needs based on operational loads and

¹²⁵ The IEA published a forecast from 2024-2026

¹²⁶ <https://digital-strategy.ec.europa.eu/en/library/energy-efficient-cloud-computing-technologies-and-policies-eco-friendly-cloud-market#:~:text=There%20is%20a%20growing%20trend,rise%20to%2012%25%20by%202025>

external conditions, data centres can afford to safely increase set temperatures. This adjustment not only reduces the energy consumed in cooling but also extends the lifespan of cooling equipment, further contributing to environmental sustainability.

Challenges

The main challenges are related to the development of algorithms that assist in the management of cloud resources, on one side, and the evolution of cooling systems to achieve better efficiency in the usage of water.

The algorithms need to consider the need for optimisation in energy consumption and prioritization in the usage of renewable energies, and down the line interact with smart grids and the overall energy systems. These algorithms rely on the availability of data to learn and enable decision making. The availability of widespread data in this area is hampered by the natural urge of companies operating in this area to regard their power usage and heating/cooling data as key confidential business data. Similar issues in the electricity generation market have been hampering developments in that area as well.

Research & Innovation Priorities

To achieve a more energy-efficient and sustainable Computing Continuum, the following R&D priorities have been identified:

- **Carbon- and energy grid-aware computing:** Adapt scheduling and orchestration algorithms to shift workloads in time and space to react to local fluctuations in renewable energy supply, as well as take advantage of excess energy supply.
- **Data centres as active participants in smart grids:** Incorporate advances in open energy data, carbon-aware grids, and forecasting of local energy mixes to optimize carbon emissions rather than just energy efficiency. The recovery of waste heat needs to be further investigated and implemented.
- **Establish a standardized set of sustainability metrics and benchmarks:** While metrics like Power Usage Effectiveness (PUE) are valuable, they fall short in capturing the efficiency of processing per unit of energy, especially in the context of a highly heterogeneous continuum.
- **Implementation of accurate energy consumption measurement in cloud-Edge systems:** In close connection with the previous priority, it is about developing the needed probes and data collectors that will gather the needed raw data that will subsequently go through an aggregation process to be shown in the shape of a series of metrics to evaluate the energy consumption and the efficiency in the management of cloud resources from a sustainability perspective. The developed probes and collectors must be easily integrable and in this sense open-source solutions should be strongly promoted to ensure wide adoption.
- **Digital twins for holistic data centre management:** Develop hybrid models and scientific/physics-based ML combining physical models, CFD and AI models e.g., to jointly control cooling, management and optimisation of the digital infrastructure for the Compute Continuum. Detailed data-driven simulation and understanding of energy flows within data centres and their integration into electricity and thermal grids, allows for the identification and implementation of strategies that can significantly enhance energy efficiency. For instance, by closely monitoring critical operational variables such as CPU

temperatures and dynamically adjusting data centre loads, it becomes possible to optimize the balance between IT performance and cooling energy consumption.

- **Adopt a whole-system approach:** It's vital to recognize that power consumption isn't confined to data centres alone but extends to the network and specific applications, necessitating benchmarking considerations based on workflows. Innovation should also target energy-efficient software for IoT devices and energy-restricted devices to minimize power consumption.
- **New cooling solutions for next-generation high-performance hardware:** For example, different liquid or even submersion cooling, to enable next-generation high-performance hardware.
- **Hardware and software co-design for specific use case applications:** For example to enable highly energy-efficient AI inference or AI training at the edge.

Potential Impact

Focusing research and innovation efforts on this subtopic will directly contribute to the EU Green Deal objectives and accelerate Europe's transition towards climate neutrality.

It will prepare the computing continuum for future smart grid technologies and significantly reduce the environmental footprint of digital infrastructure and reduce contention for energy resources with other industries.

Green Computing (including carbon-neutral computation) for all aspects of the cognitive computing continuum is a key interest for government and industrial players in Japan and South Korea.

5.2 AI-assisted operation maintenance of large-scale systems

Systems are becoming ever more complex. Whether new functionality is being bolted onto legacy systems, alternatively functionality creep, there are many systems millions of people rely on. So complex, such systems will never be completely understood, often maintained by teams of support people. A software engineering perspective is that the time it takes to find a bug, if it isn't identified when the code is written, increases by an order of magnitude when debugging, and another order of magnitude once the software is deployed or operational, see Software Engineering by Ian Sommerville.¹²⁷ Probably this is a very conservative estimate, and it is a time estimate only, not a cost estimate. DevOps is a strategy to mitigate or ameliorate the software engineering software doctrine.

Challenges

There are many reasons components will fail in the computing continuum, such as infrastructure or network failures, or an error during application deployment or operation. It is generally hard to track down the relevant information in the system logs and extract insights

¹²⁷ <https://amzn.eu/d/d5KWtm0>

from them, and there is a large gap between human-operator insights, for example in network operations centres, and system-generated logs. New solutions and tools are needed both for collecting relevant metrics and logging information across the lifetime of an application in the continuum, and for extracting relevant insights from those logs and metrics, will be needed. Not only this, automatic interpretation of those insights and proposal of remediation actions to be approved by the user can make the fixing process more efficient.

AI can also help in transforming descriptions of actions in human language into commands scripts used in the fixing process. In addition, another challenge might be the automatic deployment of temporal monitoring dashboards based on descriptions provided using natural language. AI can assist in the interpretation of the requirements and in making available a user interface to make easier for the user to follow-up the process of handling issues. The resources devoted to running these temporal dashboards can be released once it is not needed anymore. In any case AI agents can learn from each case, using such a knowledge in the next incident. All these tasks can be carried out by specialised AI agents that could be eventually combined to generate end-to-end applications.

But not only maintenance can benefit from the usage of AI agents. Daily operation can make a leap forward thanks to AI agents able to understand natural language. Prompt engineering is emerging and will make operations more accessible for a wide variety of profiles, not only purely technical ones. As an example, AI agents can interpret natural language to carry out a zero-touch deployment with certain requirements being met. In addition, AI can make such a deployment optimal meaning it is the best solution for a problem that involves a series of parameterized constraints

Research & Innovation Priorities

To deploy AI for assisting with operation maintenance of large-scale systems, the following R&D priorities have been identified:

- Develop and apply AI technologies for debugging and root cause analysis: For example, state-of-the-art AI technologies, such as large language models (LLMs) and multi-modal models, have shown an impressive ability to understand context and extract insights from semi-structured data.
- An AI and data processing system capable of synthesizing large quantities of data, including system logs, while providing a flexible human-machine-interface to query the system for relevant insights.
- Automatic interpretation of failure contexts and proposal of remediations, as well as AI-powered monitoring mechanisms such as temporal dashboards.
- A cybersecurity perspective for Security Operations Centre (SOC) teams to assess the degree of visibility they have over security incidents in the continuum for different Tactics, Techniques and Procedures (TTPs) from real-world threat scenarios. This includes identifying the different data sources in the cloud environment, training an LLM to classify which information can be extracted and verify its completeness, and aligning the results with techniques from well-established security frameworks (i.e., MITRE ATT&CK) to pinpoint how the threat monitoring capabilities of a SOC can be extended.
- AI-assisted log analysis capabilities to alleviate the workload of SOC analysts, reduce their fatigue and help them make informed and accurate decisions.

See also, section 6.5 “Operationalization of future AI systems in the Computing Continuum (GenAIOps)”.

Potential Impact

Investing in AI-assisted operation maintenance directly supports Europe's ambition for a resilient, efficient, and competitive cloud-edge ecosystem. The adoption of proactive AI to support operations and maintenance can significantly reduce downtime, operational costs, and the risk of critical failures, improving the reliability and security of European digital infrastructure. Additionally, it aligns with strengthening European competitiveness in AI technologies.

5.3 New distributed data processing paradigms for the Computing Continuum

Data is increasingly generated by devices and systems at the edges of the network. Continuous analytics of such data streams will require data management and analytics solutions that work in highly distributed environments. In use cases where these analytics workflows depend on previous data, or where the processing is distributed between edge sites, there is a need for efficient state synchronization mechanisms. Hence, there is a need for new paradigms for state management and continuous analytics in the Compute Continuum.

Both stateful and stateless architectures are widely used in applications:¹²⁸

- In stateful applications, client data is saved on the server and carried over from one session to the next. This leads to faster processing and improved performance.
- In stateless applications, client data is not preserved between sessions. Instead, they use external entities such as databases or caches to manage state.

The choice of adopting either a stateful or a stateless architecture has important implications in terms of performance and scalability.

Serverless computing simplifies the deployment and operation of applications, by abstracting away the infrastructure layer. Stateful serverless could further simplify this by abstracting away state management and synchronization, especially in a distributed computing continuum.

Challenges

There are different solutions for managing state in central cloud services, and they are offered by most market-leading cloud providers such as Amazon and Microsoft. However, there is no technical solution for managing state in a multi-provider distributed cloud-edge environments, ensuring consistency and integrity of the state. Although small research

¹²⁸ <https://aws.amazon.com/blogs/architecture/converting-stateful-application-to-stateless-using-aws-services/>

projects on novel data structures exist (e.g. on Conflict Free Data Types), there is some way to go that could provide such functionality.

Research & Innovation Priorities

To create a stateful application model for a distributed cloud-edge environment, the following R&D priorities have been identified:

- **State synchronization in the computing continuum:** Develop mechanisms for seamless state synchronization between cloud and edge, ensuring consistency and reliability.
- **Programming models to simplify application development:** Create programming models that simplify the development of stateful serverless applications for edge computing.
- **Consider resource utilization:** Synchronization and migration of state incurs additional costs in terms of for example resources and latency. Optimize resource utilization on edge devices to balance/trade-off workload distribution and energy efficiency. Multi-objective optimisation is one method to address resource utilisation.
- **Data logs consumption:** Facilitating the decision-making and the automation of operational and cyber-incident handling. See also section 4.4.

Potential Impact

This would enable future stateful continuous data and AI analytics at the edge, which include various advanced IoT and digital twin use cases and applications. A stateful application model for a multi-provider distributed cloud-edge environment will simplify the life of application developers and the management of AI and data-driven services in the continuum. It will hide the complexities of synchronising data between databases at different providers and compute clusters, which would otherwise require error-prone solutions for manually transferring data and ensuring data consistency and integrity. This could also greatly reduce network traffic and latency for such use cases. Combining the stateful and serverless computing paradigms could similarly simplify management of applications that need to manage and synchronize state between sessions or users.

5.4 Safety-critical applications in the Computing Continuum

Safety-critical applications refer to systems whose failure or malfunction could lead to significant harm to people, environment, or infrastructure. In the context of the Cognitive Computing Continuum, these include applications such as autonomous transportation, critical healthcare monitoring, industrial automation, smart grids, and emergency response systems. Ensuring the reliability, availability, and safety of these applications is critical as computing resources become more distributed across cloud and edge environments.

Next-generation intelligent systems are distributed by design, connecting multiple spatially distributed sensors, with high safety requirements and having to cope with a limited amount

of energy and often limited cooling capabilities. However, their demand for integration of state-of-the-art AI components e.g., used for perception or classification tasks, typically require high-performance hardware. High-availability cloud-edge solutions with real-time considerations are key to bridging these constraints. To achieve this, we can use an AI-powered orchestrated approach, with a reduced response time relying on performance metrics as well as on non-functional ones such as energy consumption and security, among others.

Challenges

An increase in the number of dedicated critical software functions and complexity of the underlying hardware, operating systems, and communication networks is expected. This complexity creates the need for an orchestration layer that ensures the functioning of the overall system, especially when it comes to safety-critical operations.

At the same time, it imposes new challenges in system design practices and, consequently, in gaining acceptance from the public. Due to limited compute resources on devices, it is necessary to integrate compute resources from cloud data centres and edge. It is also necessary to address privacy concerns, particularly for distributed learning models.

Research & Innovation Priorities

To effectively support the deployment of safety-critical applications in the European Cognitive Computing Continuum, focused research and innovation efforts are required in the following areas:

- **Seamless cloud-device integration:** Develop platform features that can seamlessly integrate cloud resources in onboard applications. Instead of being restricted to the data produced only within one owned system, other data sources become instantly usable using a high-availability database.
- **Real-time and deterministic computing:** Developing technologies and protocols ensuring predictable and guaranteed real-time communication and processing capabilities across distributed edge-cloud environments.
- **Fault tolerance and resilience:** Designing distributed algorithms and architectures capable of maintaining system functionality even in the presence of failures or disruptions.
- **Certification and regulatory compliance:** Innovating methodologies and tools for the certification of distributed safety-critical applications in accordance with EU safety and cybersecurity regulations.
- **Secure and trustworthy execution environments:** Developing robust security mechanisms and trusted execution environments that guarantee the integrity and isolation of critical tasks and sensitive data.
- **Risk management and predictive analytics:** Integrating advanced AI-driven predictive analytics and risk management strategies to proactively manage potential system failures or security threats. See also section 5.2 "*AI-assisted operation maintenance of large-scale systems*".

Potential Impact

Investing in these R&D priorities will have significant impacts on the EU, particularly in strengthening digital sovereignty by reducing dependency on external technology providers for critical infrastructure. Additionally, it will help ensure compliance with stringent EU regulatory frameworks (such as GDPR, NIS2, Cyber Resilience Act), enhancing public trust and acceptance. Ultimately, it will boost the competitiveness of EU industries by enabling them to deploy advanced, secure, and reliable safety-critical systems at scale, positioning Europe as a global leader in trusted digital infrastructures.

5.5 The continuum performance: cross-layer optimisation

In a multi-provider Cloud-Edge continuum, *Cloud optimisation* becomes a distributed optimisation problem in which each entity has partial information and possibilities to adjust operational parameters. A layer, often standardised, in an IT system implies the same behaviour on separated subsystems. Each layer has an identifiable functionality that designers and users depend upon. By optimising the layers, performance increase can be achieved at the cost of interoperability. In this section, network optimisation is selected somewhat as a forerunner, and as a runner up; the cloud infrastructure. Trying to optimize each layer separately will lead to suboptimal performance – this is the problem of local vs. global optimisation.

To exemplify why layering exists, a well-known system is TCP/IP, the dominant protocols of the Internet. TCP resides over IP. IP has two major versions, IPv4 and IPv6. The key insight is TCP needs minimal change to support to v4/v6 *due to protocol layering*. Again, these layers are defined as standards, in this case by the Internet Engineering Task Force, or IETF.

Challenges

The different layers of the continuum must be operated in a coordinated fashion. For example, in a data centre the distribution of workloads should be coordinated to optimize resource utilization and energy use of the cooling system.

Research & Innovation Priorities

System: Cross-layer protocol and mechanisms to coordinate the optimisation across different levels: 1) the cooling and resource utilization of the data centre infrastructure and individual servers; 2) meta-orchestration and workflow execution across machines and clusters; and 3) optimisations at the application and service levels. It is important that these coordinate network, data, and compute resources.

Network: Establishing a set of “standard” services and APIs to facilitate the optimisation across the Compute Continuum, service discovery, etc.

The “continuum” concept itself and the cross-layer approach with the existence of diverse APIs and different integration points, marketplaces, add-on catalogues and related resources have the drawback of dramatically expanding the attack surface for cyber criminals. A comprehensive approach to cybersecurity in the continuum is thus necessary so that

European citizens will make the most of its multiple benefits. Internet of Behaviours (IoB) approaches the problem by means of analytical algorithms that study the behaviour of the network and the infrastructure and can anticipate the unfolding of cyberattacks without needing actually to access the data, just looking into behavioural patterns.

Moving towards “digital shadows” and “digital twins” of the continuum, here focusing on the data flows and the integrations needed to optimize the management of network, data, and compute resources across the Compute Continuum.

Potential Impact

Hyperscalers have a big advantage, since they control their entire cloud stack from facilities to servers/hardware to service offerings. New AI-based solutions for managing the emerging Compute Continuum in a holistic way will improve resource utilization and reduce energy use across the multi-provider continuum. This will help cut down the costs of operation and provide a more viable and competitive EU alternative.

5.6 Towards a hyper-decentralized Computing Continuum

A hyper-decentralized Computing Continuum involves distributing computational resources and data management across a large number of independent nodes, including local edge nodes, local data centers, and cloud platforms. Hyper-decentralization aims to enable dynamic and peer-to-peer collaboration among these nodes without relying on a central authority or single point of failure.

This approach aligns with Europe's ambition to create a more competitive and open cloud and edge market, and reduce dependency on large, centralized non-European providers. A decentralized approach ensures greater resilience, scalability, and autonomy in digital infrastructures. This would require advances in both business and technical solutions, where functionality is pushed away from the cloud to small and individual enterprises.

Challenges

Are we ready for millions of privately-owned edge nodes as part of a future hyper-decentralized continuum? It will be difficult for centrally governed continuum federation models to deal with scenarios in which many new providers join the edge computing market, or in cases in which the market becomes hyper-fragmented.

It may be possible for individuals or small entities (e.g. under a cooperative model) to purchase and set up their own edge computing servers powered by green energy, and to make them available for hosting locally-produced data under a principle of sovereignty or even to be commercialized under a pay-per-use business model to third-parties (including local cloud service providers (CSPs) requiring to occasionally increase the capacity of their data centres with additional resources in order to deal with unexpected peaks of demand).

This will challenge our current models of service discovery, trust, and management of the application lifecycle in the continuum, as well as the scalability of many of the platforms and services that initiatives like the IPCEI-CIS are expected to produce in coming years.

Research & Innovation Priorities

To support the emergence of a hyper-decentralised computing continuum, the following research and innovation priorities need to be addressed:

- **Decentralized service discovery:** For example using distributed hash tables, and new payment methods for highly dynamic multi-provider scenarios.
- **Distributed orchestration and management:** New approaches will be needed for managing a decentralized computing continuum and application lifecycles in it. Develop robust and scalable peer-to-peer orchestration techniques capable of dynamically managing workloads, data flows, and resource allocation in a highly decentralized environment.
- **Data sovereignty and security and trust mechanisms:** Advancing distributed data security, privacy-preserving computing methods, and secure multiparty computation, ensuring data integrity and compliance across multiple jurisdictions. Explore trustworthy certification of program execution and proof of execution in decentralized computing environments.
- **Decentralized AI algorithms:** Designing AI algorithms optimized for distributed execution, including federated learning and decentralized inference methods, tailored to a large-scale, decentralized ecosystem.
- **Generative AI in a decentralized computing continuum:** Generative AI and the size of Large Language Models (LLM) pose challenges in terms of resources requirement in a constrained environment such as that of the Edge. It means that LLM architectures should be optimized and made smaller to make them suitable for Edge deployment, needing less memory, CPU, GPU resources for both training and inference. Using LLMs in the Edge can definitely contribute to hyper-decentralized continuum, to a more optimized use of resources, to energy saving and in general to a better use experience with the applications and the data being closer to the user location.
- **Business models for smaller operators to provide services:** Develop new business models and mechanisms for smaller operators (even for individuals) to offer available compute resources when available.
- **Fault tolerance and resilience:** Researching advanced fault detection, self-healing mechanisms, and redundancy strategies specifically tailored to hyper-decentralized scenarios to ensure high availability and reliability.

Potential Impact

Greater flexibility for new players to join the cloud-edge market. It could enable a more sophisticated market that could take advantage of market dynamics, such as smaller players providing edge nodes that can take advantage of local availability of distributed renewables. Opening up for smaller entities to offer up their spare compute capacity would lead to more efficient use of existing compute resources. The quick response of starting/stopping computing workloads, hyper-distributed edge nodes could help take advantage of sudden bursts of local availability of distributed renewables, that otherwise would lead to grid instability and energy curtailment, to increase the sustainability and resilience of the overall energy system.

6 Cloud-Edge for AI: Enabling and Facilitating AI Applications Across the Continuum

The Coordinated Plan on Artificial Intelligence¹²⁹ outlines the EU's strategy and approach to AI built on excellence and trust, putting people first and building strategic leadership in high-impact sectors, as elaborated in the *"White Paper on Artificial Intelligence: a European approach to excellence and trust"*¹³⁰. The importance of securing critical computing capacity in the EU can be seen in several actions that have been implemented towards this plan, for example:

- The European Chips Act and Chips JU¹³¹ (also section 3.5) which aim to address semiconductor shortages and enhance the European semiconductor industry.
- The European High Performance Computing Joint Undertaking (EuroHPC JU)¹³² (also section 3.6) initiative to develop a world class supercomputing ecosystem in Europe.
- The Testing and Experimentation Facilities (TEFs)¹³³ to support technology development of components and systems for edge AI.
- The Important Projects of Common European Interest on Next Generation Cloud Infrastructure and Services (IPCEI-CIS).¹³⁴

RISC-V has been identified as a key next-generation technology, and alternative to proprietary processor and accelerator solutions for the European computing continuum and High-Performance Computing (HPC) ecosystem.¹³⁵ The recent amendment of the EuroHPC JU Regulation to include the development and operation of AI factories,¹³⁶ further highlights the critical role of HPC in strengthening and supporting the European AI ecosystem.

In the following sub-sections, several research and development needs for the medium to long term are listed, that can be identified from these challenges and developments. They are also aligned with the AI Innovation Package¹³⁷. Although European cloud offerings span a wide spectrum of services, customers often need to collaborate with multiple providers to match the quality and breadth of services offered by leading global cloud providers. European cloud providers (including telcos) are in a unique position to be able to leverage proximity, resource optimisation, QoS evaluation, and regulatory compliance to enhance edge computing performance and address latency challenges.

¹²⁹ <https://digital-strategy.ec.europa.eu/en/policies/plan-ai>.

¹³⁰ https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en.

¹³¹ <https://digital-strategy.ec.europa.eu/en/policies/european-chips-act>, <https://www.chips-ju.europa.eu/>.

¹³² https://eurohpc-ju.europa.eu/index_en.

¹³³ <https://digital-strategy.ec.europa.eu/en/activities/testing-and-experimentation-facilities>.

¹³⁴ <https://www.bmwk.de/Redaktion/EN/Artikel/Industry/ipcei-cis.html>.

¹³⁵ https://eurohpc-ju.europa.eu/new-call-developing-hpc-ecosystem-based-risc-v-2023-02-01_en.

¹³⁶ https://eurohpc-ju.europa.eu/ai-factories-amendment-eurohpc-ju-regulation-enters-force-2024-07-09_en.

¹³⁷ https://ec.europa.eu/commission/presscorner/detail/en/ip_24_383

6.1 Portable AI applications and an open AI ecosystem

Cost and ease of use are major factors when choosing cloud service providers. For many organisations, the high costs and specialised expertise required to develop, operate, and maintain its own technology stack for developing and deploying AI applications, are a major challenge. Portability refers to the ability to move (AI) applications from one platform or service provider to another, without any or extensive adaptations to the software stack. Many have been drawn into the cloud service ecosystems of the global hyperscalers, due to the convenience provided by the broad and integrated service ecosystem offered by them. These service ecosystems make it easy to develop and deploy data-driven applications and AI but also lead to vendor lock-in since it is prohibitively expensive to move a system from one provider to another. This applies both to private and public sector, and in particular SMEs that can use free credits to quickly create systems based on cloud technologies and AI.

Challenges

One major challenge for Europe is that European datacentre providers lack the resources required to offer such broad and complete AI service ecosystems. Different use cases and applications will require different types of AI models, and different ways of training and deploying them. We can expect increased hardware heterogeneity as new AI-optimised hardware becomes available for different parts of the spectrum, for example supercomputers and devices. AI technology stacks should not be restricted to Cuda and x86. One approach is porting ML libraries, and another one is creating ML compilers – both need to be supported to facilitate a “technology neutral” computing continuum. There is a need to standardize models/formats that are used for communication, so they can execute efficiently on different platforms. Public funding can help develop an AI service ecosystem by building and demonstrating it on open standards and open-source components.

AI workloads can be resource-intensive and require careful orchestration to ensure efficient use of cloud resources. EuroHPC provides high-performance computing (HPC) resources for AI development, with plans to expand the capacity, but traditional HPC is sometimes considered too different or too complicated for some companies. For example, file management, such as moving files from the cloud to the supercomputer, can take time during (AI) model training.

Tools like the meta-orchestration platform ColonyOS¹³⁸ can be beneficial as they create a level of abstraction, for example simplifying file management and deploying cross-platform applications.

Research & Innovation Priorities

Develop an AI service ecosystem built on open standards and open-source components, including:

- Efficient and scalable data management from IoT devices in the continuum.

¹³⁸ <https://colonyos.io/>

- Inference in the continuum, including i) where to perform inference in the most energy-efficient way, ii) capabilities for running inference, which connects to virtualization topics (section 6.4), iii) compatibility with a distributed network of compute services.
- Training in the continuum, for example integrating HPC resources or solutions for distributed/federated learning use case.
- MLOps solutions in the continuum.

Develop technology-neutral and efficient execution on different platforms:

- Software-hardware co-design and solutions for optimising the AI stack as new hardware architectures become available, in collaboration with efforts in 6.3.
- Develop AI compilers, e.g. from Python to native, to facilitate cross-platform by design.
- Develop more energy-efficient AI solutions through software-hardware co-design.
- Efficient resource allocation for AI workloads in the HPC integrated continuum: Develop strategies for dynamic resource allocation based on workload requirements to optimize resource usage in the Compute Continuum for AI.

Lastly, there is a need to establish open standards and APIs for an AI ecosystem in the computing continuum by standardizing models and formats used for communication, so they can execute seamlessly on different platforms.

Potential Impact

An open AI service ecosystem can enable a seamlessly interconnected multi-provider computing continuum, facilitate collaboration between different parts of the value chain, and increase “technology neutrality” in the continuum. Open standards and APIs can create new markets and opportunities. For example, it would be possible to have brokers that provide AI services that are portable in the computing continuum.

Investing in technology-neutral technologies will address hardware heterogeneity and make the software stack and skills more portable and robust, avoiding high reliance on vendor-specific hardware-optimized software, such as cuDNN.¹³⁹

See also section 6.2 and 6.5.

6.2 Integrating HPC in the Cognitive Computing Continuum

With a budget of EUR 7 billion for the period 2021-2027, the EuroHPC JU coordinates efforts and pools resources across Europe to develop and maintain a “world-leading federated, secure and hyper-connected supercomputing, quantum computing, service and data infrastructure ecosystem,” to boost scientific excellence and industrial strength in Europe.¹⁴⁰ The *AI Innovation Package* further establishes the important role of the EuroHPC supercomputer ecosystem in

¹³⁹ <https://developer.nvidia.com/cudnn>

¹⁴⁰ https://eurohpc-ju.europa.eu/about/discover-eurohpc-ju_en

the future development of general-purpose AI in Europe, through the development and operation of AI Factories and supercomputers optimised for AI.¹⁴¹ As an example, the flagship initiative Destination Earth (DestinE)¹⁴² will exploit the distributed high-performance computing provided by EuroHPC to develop a highly accurate digital model of the Earth on a global system. DestinE will further consolidate access to valuable sources of data across Europe and represents a key component in the European strategy for data and will contribute to the twin transition, green and digital.

In May 2024, the United Kingdom joined the EuroHPC Joint Undertaking, becoming the 35th participating state. This inclusion allows UK researchers and scientists to apply for Horizon Europe-funded EuroHPC JU Research & Innovation calls, enhancing collaboration and resource sharing across Europe. The UK's expertise in supercomputing will contribute significantly to Europe's objective of becoming a global leader in HPC and quantum computing. Additionally, the Hanami Project, a collaboration between Europe and Japan, was established. It aims to promote scientific projects involving both European and Japanese institutes. This project facilitates access to supercomputers in both regions, tackling exascale areas and beyond. Hanami focuses on priority domains such as climate and weather modelling, biomedical sciences, and materials science, fostering a sustainable collaboration between Europe and Japan.

Challenges

It is difficult for industry to adopt EuroHPC JU supercomputers due to various challenges associated with HPC environments, such as differences in software stacks compared to cloud services. For example, there are significant differences in IT environments (e.g., Kubernetes vs. Slurm), and processes for deploying multi-cloud federations involving cloud, edge, and HPC resources. Additionally, the heterogeneity of HPC constitutes a challenge. For instance, the EuroHPC system LUMI, which uses AMD processors, uses the ROCm open-source software stack, while NVIDIA has its proprietary CUDA stack. However, machine learning libraries are not as mature on AMD GPUs compared to NVIDIA. Development and improvement of these open-source stacks are required to help bridge the gap between different HPC systems. Over time, as EuroHPC adopts new processors manufactures in Europe based on ARM and RISC-V, there will be a need for optimizing software for different hardware architectures. For example, HPC systems can have compatibility issues with standard Machine Learning (ML) libraries due to differences in their software stacks, which can hinder the development and deployment of AI services in the cloud.

Research & Innovation Priorities

Reduce the skill gap for HPC by providing concrete scenarios and guidelines to help industries understand and overcome these challenges. There is a particular need to bridge the gap between EuroHPC supercomputers and EU cloud ecosystems, embracing a hybrid cloud paradigm:

- Promote the development of an open-source ecosystem to facilitate the integration of HPC and Cloud services.

¹⁴¹ <https://digital-strategy.ec.europa.eu/en/factpages/ai-innovation-package>

¹⁴² <https://destination-earth.eu/>

- Encourage the use of and further develop open-source software stacks to enable highly performance training of AI models in EuroHPC supercomputers.
- Simpler user interfaces and better portability of workflows would make it easier for companies to adopt HPC and integrate it with their cloud deployments.
- Support projects like EPICURE, which aims to enhance user support by establishing and operating a distributed but coordinated European HPC application support service. EPICURE encourages the best possible uptake of HPC systems by European researchers, fostering a collaborative environment for HPC and AI development.¹⁴³
- Optimize strategic open-source software, including system software and libraries for AI, for EuroHPC architectures to enable full use of emerging and installed architectures.

Potential Impact

Foster the industrial uptake of EuroHPC and AI. If implemented, this solution would make it easier for European startups and industries to adopt HPC and leverage the EuroHPC initiative, potentially leading to more efficient use of resources, significant cost savings, improved competitiveness of EU industry in AI technologies, and advancement towards the Computing Continuum. These challenges are also relevant for cloud computing and the introduction of new hardware architectures, synergizing with sections 6.3 and 6.4.

6.3 RISC-V and new processing initiatives in the Cognitive Computing Continuum

Over the past decade and a half there has been significant investment in implementing European processors targeting both HPC and embedded and IoT applications. The EU goal is the production of cutting-edge and sustainable semiconductors in Europe, with at least 20% of world production in value by 2030, including manufacturing capacities below 5nm nodes, aiming at 2nm, with an aim to improve energy efficiency by a factor 10. To achieve this the RISC-V ISA plays a central role in EUs strategy. RISC-V, as an open specification, enables Europe to develop processors and domain-specific accelerators (including AI) benefitting from local and global ecosystems but without royalties or intellectual property dependencies. Together with trends such as chiplets, that increase yield and lower manufacturing costs, this lowers the barrier of entry to the microprocessor chip market for European companies and can drive demand for the manufacturing capabilities pursued in the European Chips Act. The European efforts, spanning the Compute Continuum, expose a strong European commitment to a joint European ecosystem based on RISC-V.

Challenges

Today's Compute Continuum is dominated by Intel's x86 in cloud/HPC and by ARM's 32- and 64-bit Instruction Set Architectures in edge and embedded domains, and most tools and techniques used in these domains are specialised for this architecture. ARM has recently made the push towards cloud and HPC which has resulted in significant efforts to port applications

¹⁴³ <https://epicure-hpc.eu/>

such as container managers, hypervisors, AI frameworks, as well as OS drivers for hardware. For broad adaption RISC-V also needs to address these software aspects as outlined in previous roadmaps and identified in the RISC-V community through collaborations such as RISE¹⁴⁴. Recent results point to a performance gap to established non-European proprietary instruction set architectures¹⁴⁵ which must be closed to enable wide-spread adoption of RISC-V. Furthermore, the openness of RISC-V leads to faster innovation cycles which opens the question of how to manage and abstract increasing diversity of RISC-V computing hardware from end-users (and expose when necessary).¹⁴⁶

Research & Innovation Priorities

Software tools and technologies to handle the unique challenges and opportunities with fast-innovation open instruction set architectures should be researched to abstract and manage expected increase in hardware diversity. Additional development needs are required to port software tools and technologies relevant to the Compute Continuum to RISC-V and close the performance gap to systems deployed on established platforms, these include above identified known challenges as well as other challenges identified in this chapter, including the challenges in section 6.2 and 6.4. The efforts rely on retained momentum in ongoing RISC-V hardware efforts, including for AI, supported by EuroHPC JU and Chips JU to drive demand and generate necessary spillover effects in European RISC-V hardware. Industry sectors across the Compute Continuum may have other needs than e.g., HPC, automotive, and space. Therefore, key use cases from these cognitive computing continuum industry sectors, and their supply chains, should be identified and used to drive additional software and hardware research and development needs for the European ecosystem, complementing ongoing Chips and EuroHPC initiatives.

Potential Impact

European RISC-V hardware and software ecosystem can be extended to encompass the entire cognitive Compute Continuum, extending the European ecosystem and leveraging technology spillover effects from R&D in HPC (EuroHPC JU), automotive, and space (Chips JU), while at the same time strengthening the ability to adopt RISC-V across more European industrial sectors. Ongoing European efforts in selected industries serve as key locomotives in this strategy, but as outlined in e.g., 6.2, there is additional potential for the European RISC-V ecosystem in making it adoptable by a larger share of the economy. Efforts in this domain synergize with 6.2 and 6.4. Traditional microprocessors, including those based on RISC-V, remain key for most compute areas of most sectors, but can when deployed in tandem with longer-term disruptive technologies like neuromorphic and/or quantum computing (section 9) significantly improve performance and energy-efficiency in certain domains.

¹⁴⁴ RISE: RISC-V Software Ecosystem, Linux Foundation, <https://riseproject.dev/>

¹⁴⁵ Francesco Lumpp, et al, "On the Containerization and Orchestration of RISC-V architectures for Edge-Cloud computing," ESAAM'23, 2023.

Robert Balas, et al, "CV32RT: Enabling Fast Interrupt and Context Switching for RISC-V Microcontrollers," IEEE Transactions on VLSI Systems, 2024.

Andrea Bartolini, et al, "Monte Cimone: Paving the Road for the First Generation of RISC-V High-Performance Computers," SOCC'22, 2022.

¹⁴⁶ <https://digital-strategy.ec.europa.eu/en/library/recommendations-and-roadmap-european-sovereignty-open-source-hardware-software-and-risc-v>

6.4 A middleware toolkit for the Computing Continuum

The Compute Continuum will almost certainly be a heterogeneous set of hardware and software systems. Therefore, running applications on different architectures, such as ARM, Intel (x86), RISC-V and differing operating systems makes it hard to find a one size fits all for software developers coding in the continuum. In large organizations even specialised teams e.g., IOS, Android are needed. Tools applying the Infrastructure as Code (IaC) approach, such as Ansible, may help, but are still often too error prone. There are several ongoing open initiatives to address this problem, including the open W3C standard called WebAssembly¹⁴⁷ and the open LLVM¹⁴⁸ project, a collection of modular and reusable compiler and toolchain technologies, with its associated Multi-Level Intermediate Representation (MLIR)¹⁴⁹ project. These initiatives enable producing code from (m)any high-level languages. The expression “write once, run everywhere” typically formulates this principle.

Challenges

Closing the performance gap with native binaries. There is about 1.2 performance penalty compared to natively compiled code. For full utilization of MLIR domain specific languages offer higher levels of abstractions that facilitate multi-level code generation across heterogeneous targets. AI development is strongly driven by domain-specific languages, but this is less pronounced in software development in many industry domains.

Europe is strong in compiler and programming language technology. A strong startup ecosystem is needed around these technologies, such as WebAssembly, as well as large company adoption. Promote usage in series such as the Dagstuhl seminars.¹⁵⁰

Research & Innovation Priorities

- A modern middleware, and support for lightweight virtualization/container technologies: Use compiler technology to produce machine-like code for any architecture. Support for lightweight virtualization technologies on devices, such as WebAssembly or Unikernels, is needed.
- Heterogeneous Computing Support: Improve support for heterogeneous computing environments to ensure efficient AI workload execution across different hardware architectures, and in particular interoperability with RISC-V.
- Open and community-driven development: Promote a community-driven approach to foster innovation and contributions to open ecosystems, such as the LLVM and MLIR ecosystems.
- Standardization and collaboration: Encourage standardization and collaboration among industry stakeholders to accelerate the development of open compiler infrastructures.

¹⁴⁷ <https://www.w3.org/groups/wg/wasm/>, <https://webassembly.org/>.

¹⁴⁸ <https://llvm.org/>.

¹⁴⁹ <https://mlir.llvm.org/>.

¹⁵⁰ <https://www.dagstuhl.de/en/seminars/dagstuhl-seminars>

Potential Impact

The ambition of the Computing Continuum will feature high platform heterogeneity. A middleware toolkit can help harmonize the Compute Continuum, using open standards, according to the principle “write once, run everywhere”. Factory 4.0 and 5.0 as well as systems with many IoT devices are ideal candidates.

6.5 Operationalization of future AI systems in the Computing Continuum (GenAIOps)

The term GenAIOps refers to successfully implementing, overseeing, and refining AI applications in a heterogeneous and constrained networked computing environment.¹⁵¹ The term encompasses best practices from other operational best practices such as DataOps (Data Operations), LLMOps (Large Language Models Operations), and DevOps¹⁵².

It is common knowledge that the implementation of AI techniques in the computing continuum is not only desirable but also necessary to support future applications and continue to evolve. Certain requirements, like power consumption or storage, make the application of AI techniques particularly difficult when discussing certain computing continuum components, like IoT devices. Moreover, the deployment is made even more difficult by the inherent heterogeneity of the computing. Mechanisms to function and improve the AI implementation are essential in this situation.¹⁵³ For instance, the minimization of the various elements that comprise the computing continuum can facilitate the operationalization of Generative AI applications. Techniques such as the reduction of the data transmitted between the various computational layers, and the minimization of the operational resources of the models facilitate the operationalization of such models.

Challenges

Deploying emerging AI learning technologies in the continuum introduces new challenges. Specifically, Large Language Models (LLMs) and Foundation Models (FMs) are very large and costly in terms of resources such as power, time, and storage, and therefore challenges the lower/smaller end of the continuum. Other issues for the continuum are heterogeneity and full-stack dependencies. Infrastructure needs for FM training and inference, i.e., pre-trained on vast amounts of unstructured data to learn complex concepts, need to be addressed.¹⁵⁴
¹⁵⁵

Research & Innovation Priorities

- *Dynamic Orchestration*: Implement comprehensive OS solutions to simplify the deployment of future AI systems in the continuum addressing the specific needs of future

¹⁵¹ *Continuum AI: Integrating Foundational AI Agents with the Cognitive Computing Continuum*

¹⁵² <https://www.karini.ai/blogs/navigating-genaiops-in-enterprises>

¹⁵³ *The Operationalization of AI*

¹⁵⁴ <https://dl.acm.org/doi/full/10.1145/3625289>

¹⁵⁵ *Autonomy and Intelligence in the Computing Continuum: Challenges, Enablers, and Future Directions for Orchestration*

AI models, such as FMs and LLMs plus the combination of classic ML models. A computational aspect is to use compression techniques to reduce computational cost, memory footprint and energy consumption. An example is to enable edge-based deployment such as using TinyMLOps.

- *Computational strategies for AIOps:* Training and deploying generative AI models can be resource-intensive, requiring significant computational power and storage. A computational strategy to address this challenge might imply the usage of compression techniques to reduce computational cost, memory footprint and energy consumption. An example is to enable edge-based deployment such as using TinyMLOps.
- *Security and privacy:* Security and privacy are critical concerns when it comes to generative AI systems. Given the widespread and diverse nature of computational resources and the intensive use of ubiquitous data in generative AI systems, security and privacy challenges must be addressed from a holistic perspective. Issues such as data poisoning, output manipulation, an increased attack surface, and limited resources for robust security measures necessitate novel approaches to secure the entire supply chain of generative AI systems.

Potential Impact

European competitive advantage and independence for upcoming AI solutions deployed in the continuum.

6.6 Federated computation for Foundation Models

The radical shift of Federated computation for Foundation Models training could lead to more specialised and efficient AI architectures that are capable of handling the complex demands of training large-scale neural networks while preserving privacy and leveraging distributed data sources. This shift is part of the broader trend towards more powerful and versatile AI systems that can be adapted for a multitude of business and consumer applications.

Federated computation could play a significant role in the future of AI training by enabling more efficient and privacy-preserving methods to train foundational models. The creation of foundation models, which are a subset of large language models, has changed the methodology of how AI can be created. These models can be adapted for a wide variety of use cases with much higher productivity than before. Federated computation could further enhance this by allowing for distributed training across various devices and data centres, potentially leading to more robust and widely applicable AI models. Federation computation also has the potential to comply with the stringent requirements of existing regulations outlined in section 0. In addition, it promotes the potential of efficient use of resources by leveraging the computational power of the infrastructural devices in the edge of the network, which can reduce the load on other servers in the architecture and the network. Finally, the possibility of keeping the data on local devices significantly reduces the risk of data breaches

and unauthorized access. Further information is found e.g. in the reports by Woisetschläger et.al and Zhuang et.al.^{156,157}

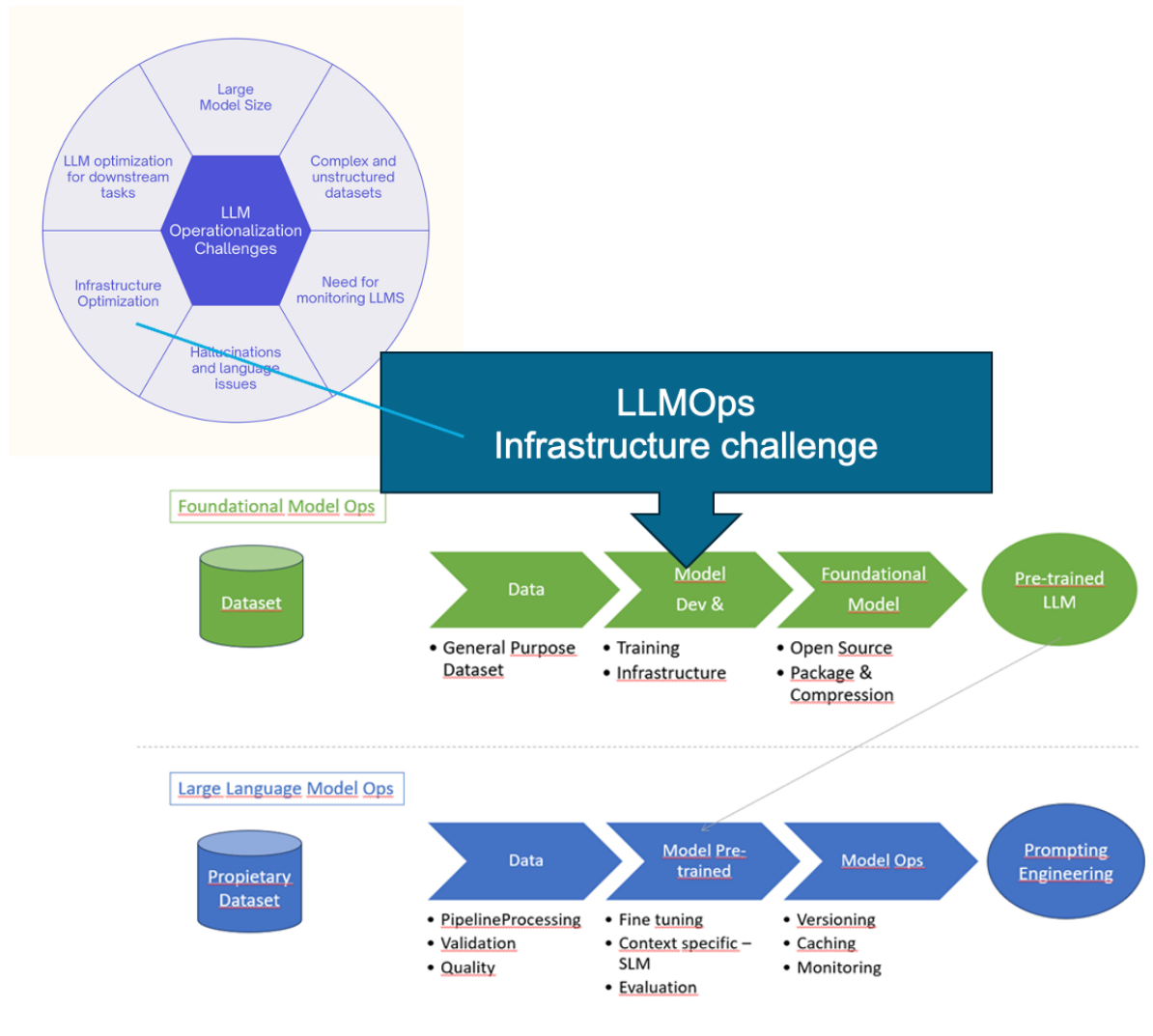


Figure 6 Large Language Models Operations (LLMOps)

Challenges

Most LLMs are highly demanding and mainly use high performance computing grade equipment. The training of the foundational models (FM) can take advantage of the recent advances on GPUs and use the computation offered by the Open continuum in what is called “Federated Computation”. Implementing federated computation in LLMs is a complex task that requires careful consideration of technical, ethical, and legal aspects. It’s a multi-

¹⁵⁶ Woisetschläger, H., Isenko, A., Wang, S., Mayer, R., & Jacobsen, H. A. (2024). A survey on efficient federated learning methods for foundation model training. *arXiv preprint arXiv:2401.04472*.

¹⁵⁷ Zhuang, W., Chen, C., & Lyu, L. (2023). When foundation model meets federated learning: Motivations, challenges, and future directions. *arXiv preprint arXiv:2306.15546*.

disciplinary effort that involves advancements in machine learning, data privacy, and distributed systems. The high complexity of foundational models and the resource heterogeneity found in federated computation makes the process even more cumbersome. In addition, communication efficiency and scalability issues must be addressed, as well as data privacy and security concerns.

Research & Innovation Priorities

Implementing federated computation for large language models (LLMs) requires several key considerations:

- The implementation of federated computing for Large Language Models (LLMs) requires establishing a distributed network infrastructure with reliable communication protocols.
- Prioritizing data privacy through techniques like differential privacy or secure multi-party computation.
- Adapting LLM designs to support federated computation may involve modifying them for decentralized data sources and partial updates.

Potential Impact

Developing efficient federated computation algorithms capable of handling LLM complexity is imperative for effective model aggregation. Additionally, optimizing resources ensures scalability and efficiency, while rigorous evaluation and testing across diverse data distributions and network conditions ensure performance. Encouraging user participation through transparency and trust-building measures and continuous monitoring for issues like data drift are also vital. Fostering collaboration among academia, industry, and regulatory bodies is essential for standardizing federated computation methodologies and advancing LLMs.

6.7 Large-scale testbeds for AI services in the Computing Continuum

US actors dominate the development of AI, possibly with non-EU values. There is a need for large-scale testbeds that can simulate real-world conditions for testing AI services, pipelines, and workflows in the Computing Continuum. There are also many emerging technologies and computing platforms that various technology and service providers must adopt or integrate with. Without such test environments, it's challenging to validate and optimize these services for use in operational conditions.

Challenges

Setting up large-scale testbeds between multiple countries and providers requires coordination between many parties. The testbed needs to be sufficiently rich to provide environments where information can be shared in real-world conditions. The testbed also need to integrate emerging technologies such as 5G/6G. These areas tie into many common technical challenges, as well as challenges stemming from regulations and other non-technical barriers.

There is demand for an International Testbed for “Cross-border Data flows”, for example with Japan and South Korea, considering the “Cross-border Data flow deal” signed between EU and Japan.

Research & Innovation Priorities

To support accelerated deployment of AI services on EU-scale, and prepare technology and service providers for the single European digital market, large-scale testbeds are needed:

- Develop large-scale testbeds that can accurately simulate real-world conditions, enabling thorough testing and optimisation of AI services, e.g. AI Factories.
- Real life 5G/6G-integrated Compute Continuum testbeds for projects developing applications using the low latency introduced by edge computing in 5G networks, and for projects optimizing the digital infrastructure, software architectures, and operations and management of such networks of computation and communication.
- Managing cross-border data flows.

Potential Impact

Prepare technology and service providers for the single European digital market and emerging technologies on the market. Large-scale testbeds are needed to accelerate EU-scale deployment of AI services, and to provide technology and service providers with a testing ground for new technologies (such as 5G) before they are deployed at scale.

6.8 Data privacy and security in AI services

Data-driven insights can contribute to vital decisions in many domains (e.g. crisis management, predictive maintenance, mobility, public safety, and cybersecurity). However, obtaining the trust of decision makers to exploit data-driven insight is still a pending issue due to i) the data and their fluctuating quality and volumetry and ii) the finality of big-data processing not necessarily suited to decision-maker comprehension.

Challenges

AI services often require access to sensitive data, making data privacy and security a significant challenge. This is especially true in multi-provider cloud federations where data might be stored and processed in different jurisdictions with varying data protection laws. Implementing this solution would increase trust in AI services and could lead to wider adoption, particularly in sectors handling sensitive data.

Research & Innovation Priorities

Implement robust data protection measures and comply with regulations like the AI Act to ensure data privacy and security in AI services:

- The development of advanced privacy toolkits. Solutions that can focus on the encryption of data records by exploring attribute-based encryption and proxy re-encryption, symmetric encryption, and searchable encryption. Additionally, could consider the

combination of Searchable Encryption and Distributed Ledger Technology (DLT) platforms.

- End-to-end secure privacy-preserving AI pipelines leveraging confidential computing technologies.
- Ensure privacy in the information exchange and certification processes. The application of Zero Knowledge Proof technology and Layer 2 solutions like ZK-EVM (Zero Knowledge Ethereum Virtual Machine) could help too as a part of the solution.

Potential Impact

As AI services are deployed in the multi-provider Computing Continuum, and data is shared between stakeholders, it is critical to ensure the privacy and security across their whole lifecycle.

6.9 Confidential computing

Confidential Computing comes into scene to solve the persistent problem related to the security of data when they are outsourced to the cloud. With the wide acceptance and adoption of the Continuum concept including all its advantages, the security problem has been amplified and extended to all tiers of the Continuum. Confidential Computing provides security to the data-in-use by means of operations performed in a hardware-based Trusted Execution Environment (TEE). Executing code inside a TEE makes it both invisible and inaccessible to any unauthorized party, including the operating system itself.

Extending this reasoning, the concept of Confidential Virtual Machine (CVM) was born. CVMs are characterized by the fact that they are run entirely within a TEE, what makes them secure. Secure and confidential computing supported all the way down to the hardware level could enable new applications with stricter data privacy requirements. Parts of this solution exist today, such as secure enclaves and Trusted Execution Environments, in a commercial context, for example solutions from Intel, SGX and TDX, AMD, SEV-SNP and IBM.

Challenges

Confidential VMs are part of the more general concept of confidential virtualization. Despite the notable progress in the subject, there are still notable challenges posed by heterogeneous platforms that work on top of extremely dynamic multi-provider infrastructures that are geographically distributed, present constraints in the use of computation resources and are connected to public networks

The various CPUs and solutions require different toolchains, as compiler support is needed to mark code parts that should execute in the enclave. A turnkey confidential computing functionality not dominated by a large, single player does not really exist. This fragmentation and lack of standardization is similar with, for example, HPC.

Research & Innovation Priorities

The R&I efforts should go in the direction of achieving confidential clouds to eventually have a confidential continuum well protected end-to-end against external threats, where data is secure at rest, in transit and in use, while at the same time satisfying the requirement of being

multi-provider and avoiding vendor lock-in. Open source management, automation, and orchestration of resources with the support of AI need to be adapted to be an effective part of a confidential environment. Efforts need to focus on the following priorities:

- Portability of confidential VMs across the Continuum.
- Open architectures for the hardware part, see further regarding the enclaves at RISC-V¹⁵⁸ and section 6.3.
- Increased support for different hardware solutions in virtualization/container technologies.
- Multi-tenancy in data centres allows access to parts of datasets, often Kerberos based.
- Securing and trusting the endpoints, think “TLS++” as a concept, in which one can trust the endpoints of the connection, the keys, and the end architecture.

Potential Impact

Confidential computing is enormous and until now unresolved. People, companies, organisations, even countries could gain value by sharing data, or parts thereof, but they do not or cannot. Therefore, it is a huge opportunity for Europe to take the initiative and the leadership in the race with hyperscalers to achieve a confidential continuum that protects sensitive data and communications. In consequence, confidential computing becomes a fundamental characteristic for the future European Technology Stack meant to lay the foundation of the European Digital Sovereignty. Additionally, it can boost the European Technology Ecosystem based on open source solutions and increase European digital competitiveness

6.10 Data spaces for AI in the Computing Continuum

Within the EU Artificial Intelligence Act (AI Act), GenAI is considered as a type of general-purpose AI model (GPAI)¹⁵⁹, and it is described as a highly capable model trained on large and diverse datasets and designed to perform a wide variety of tasks with a significant generality, allowing them to be integrated in a variety of downstream systems.¹⁶⁰ Data Spaces are “Interoperable frameworks, based on common governance principles, standards, practices and enabling services, that enable trusted data transactions between participants. Data Space enabling services are implemented by one or more infrastructures”.¹⁶¹ The common European data spaces were introduced in the EU data strategy¹⁶² and are also referenced in the Data

¹⁵⁸ <https://riscv.org/>

¹⁵⁹ AI Act recital 99.

¹⁶⁰ For the legal definition of ‘General-Purpose AI model’ see AI Act Art 3 (65). Also see discussion in Fernández-Llorca, D., Gómez, E., Sánchez, I., & Mazzini, G. (2024). An interdisciplinary account of the terminological choices by EU policymakers ahead of the final agreement on the AI Act: AI system, general purpose AI system, foundation model, and generative AI. *Artificial Intelligence and Law*, 1-14.

¹⁶¹ Why data spaces? A business and user’s perspective. DSSC, 2023.

¹⁶² European Commission (2020), *Communication on A European strategy for data*, COM(2020) 66 final.

Governance Act¹⁶³ and the Data Act¹⁶⁴ fostering qualities or characteristics related to interoperability, collaboration and the creation of synergies among data spaces, as well as promoting European values and rules, and the extended purpose beyond business value, Both GenAI and common European data spaces are interrelated components of the EU digital strategy to foster a robust and secure digital economy and promote innovation.¹⁶⁵

Challenges

Artificial Intelligence (AI) systems require robust and diverse datasets to learn, adapt, and provide accurate outputs. These datasets can be produced by different stakeholders or companies and stored along the cloud continuum from edge devices to fog or cloud environments. Each data producer can have different strategies for data sharing, including monetization of data, constraints on the companies allowed to use the data and the expected use of the data. Furthermore, trust, data sovereignty and data quality are paramount to create a data governance framework able to build AI solutions using data spread all along the cloud continuum. Data Spaces are crucial in this context as they provide the infrastructure for secure data sharing and management, which is essential for the development and deployment of AI. Data space technology still lacks the maturity level needed for seamless deployment in a cloud continuum environment. However, current initiatives in the EU, like Gaia-X, IDS and the SIMPL project among others are designing and developing the standards, tools and frameworks needed to facilitate the creation and operation of data spaces.

Research & Innovation Priorities

Data spaces hold significant promise in enabling diverse opportunities within Generative AI (GenAI). However, to fully capitalize on this potential, several critical challenges must be addressed. These challenges, inherent in the current design and implementation of data spaces, can hinder the seamless and efficient integration of data spaces with GenAI technologies. The following challenges outline the key barriers and their implications for realizing the synergy between data spaces and GenAI:

- **Interoperability and simplicity:** AI systems must effectively work with data from diverse domains and sources. To achieve this, Data Spaces must adopt interoperability standards that enable seamless integration and analysis of data across various platforms and environments. However, the current state of Data Spaces lacks simplified and harmonized access to cross-sectoral and federated data. For stakeholders such as startups, SMEs, and initiatives contributing to European AI infrastructure (e.g., AI factories), connecting to each individual data space to meet training data needs is impractical. Ad-hoc solutions to data access reduce efficiency, highlighting the urgent need for a unified approach to accessing these diverse and distributed data resources.

¹⁶³ Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724.

¹⁶⁴ Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828.

¹⁶⁵ The new "Generative AI and Data Spaces" white paper of the Strategic Stakeholder Forum is now available - News - Data Spaces Support Centre

- **Data Representation, Diversity, and Quality in GenAI Development:** For Generative AI (GenAI) models to perform effectively across a variety of scenarios—both within specific contexts and while generalizing—the datasets accessible via data spaces must be diverse, high-quality, balanced, and represent a wide range of domains, topics, contexts, and modalities. Achieving this requires rigorous data cleaning, validation processes, and continuous monitoring to maintain the integrity and usefulness of the datasets. High-quality data must capture a broad spectrum of variations, including different modalities and linguistic differences, ensuring models can adapt to diverse inputs and scenarios. The development and management of data spaces face several challenges:
 - **Scaling Adoption:** Encouraging broader adoption of common European data spaces is critical to enhance access to diverse data sources and expand possibilities for training GenAI models.
 - **Complexities in Design and Management:** Addressing design and management issues is crucial to meet the requirements of GenAI. This involves creating design methodologies and components that cater to GenAI-specific needs, such as data models, provenance, traceability, data quality, governance, and ethical considerations. Failing to address these complexities could lead to inefficiencies, inconsistencies, and obstacles in leveraging data spaces effectively.
- **The Role of Data Spaces in the Compute Continuum:** Data spaces are distributed across the computing continuum, where data may be produced and processed at multiple levels by entities with differing data policies. To support GenAI effectively, data space solutions must accommodate this diversity and enable seamless operation across the Compute Continuum. The key considerations include:
 - **Data Availability:** AI models require access to large volumes of data to train effectively. Data spaces facilitate the aggregation of data from various sources, making it readily available for AI processing, including methods derived from high-performance computing (HPC) workloads.
 - **Data Governance, Security and Privacy:** Security and Privacy: Protecting sensitive data is critical, especially when used in GenAI applications. Data Spaces provide secure environments that protect data privacy and ensure that GenAI systems are not exposed to data breaches or unauthorized access.
 - **Data Quality and Integrity:** The performance of AI models is directly linked to the quality of data they are trained on. By leveraging data spaces to maintain high-quality standards, it is possible to ensure that training data is accurate, complete, and reliable, thereby improving AI outputs.

Potential Impact

Data spaces are a key enabler for data sharing in Europe, and requires integrating data spaces in a multi-provider computing continuum. European data spaces can help ensure that GenAI models are built on a foundation of diverse, high-quality, and well-governed data. This requires overcoming challenges in scaling adoption, managing complexities in data space design, and addressing issues related to data quality and governance. Data spaces play a pivotal role in supporting GenAI development by providing robust, ethical, and efficient frameworks across the Computing Continuum.

7 Telco Cloud-Edge: Telco as One of the Main Tenants and Infrastructure Providers

Europe has enjoyed a long-standing history as a global leader in the telco industry, but there are potential risks to this position. As the telco industry moves toward a landscape with increasing integration of cloud-native frameworks and open paradigms, such as open radio access networks (initiatives include O-RAN¹⁶⁶ and the Telecom Infra Project OpenRAN project group¹⁶⁷), the traditional telco industry in Europe faces competition from non-EU cloud and tech industry. This marks a pivotal shift towards more dynamic and interoperable infrastructure. The transition is crucial not only for maintaining Europe's leadership in the global telco industry but also for enhancing its competitive edge in an increasingly digital world.

Furthermore, Europe's economy faces a considerable risk due to the EU's slower progress in expanding fibre coverage reaching 56% of European households¹⁶⁸ and rolling out standalone 5G networks covering 81% of the population¹⁶⁹, but lagging behind in 5G standalone deployment, particularly in contrast to the advancements seen in regions such as South Korea and China. This not only hampers the industry's ability to remain competitive but also reduces the possibilities for collaboration, underscoring the urgency for Europe to accelerate its cloud and network infrastructure developments to support the capabilities of the telco industry.

These challenges, and the convergence of connectivity and computing, are outlined and discussed in the white paper *"How to master Europe's digital infrastructure needs?"*¹⁷⁰ which proposes the creation of the *"Connected Collaborative Computing Network"*, or *"3C Network"*. Several research and development needs for the medium to long term can be identified from these challenges and developments:

- **Open reference architectures** that include standards for open cloud and edge infrastructure, AI-based predictive maintenance of datacentres and cloud and edge infrastructure, advanced networking, synchronization of time and requirements across providers, automated operations, and APIs for bare-metal as a service.
- **Optimized infrastructure design** focusing on innovative data centre designs that include advanced simulation and digital twin technologies for predictive capabilities. This is done alongside improvements in security, accessibility of physical infrastructure, open hardware and operating systems, operation management & monitoring (including network orchestration with multi-cloud orchestration), network functions at the edge, and connectivity between cloud and edge at scale.
- **Scalable and secure edge deployments** by establishing robust next generation edge infrastructure (forward-looking data centre facilities, servers, storage, mobile and

¹⁶⁶ <https://www.o-ran.org>

¹⁶⁷ <https://telecominfraproject.com/openran/>

¹⁶⁸ https://www.europarl.europa.eu/RegData/etudes/BRIE/2024/762298/EPRS_BRI%282024%29762298_EN.pdf

¹⁶⁹ https://5gobservatory.eu/wp-content/uploads/2023/10/BR-19_October-2023_Final-clean.pdf

¹⁷⁰ <https://digital-strategy.ec.europa.eu/en/library/white-paper-how-master-europes-digital-infrastructure-needs>.

fixed/stationary interconnections) to manage the technological complexity of the meshed continuum.

- **Forward-Looking infrastructure for building next-generation data centre** facilities and network connections that support the complex requirements of a fully integrated Cloud-Edge continuum. Develop and set up physical and logical linking of networks including integrated smart network services for the Compute Continuum to ensure seamless connectivity between cloud services and edge devices across the entire network at scale. Looking further ahead, this can also include satellite communication and other non-terrestrial networks as discussed in more detail in Section 9.2.

It is also important to take other strategic documents into account such as the IPCEI-CIS Value Chain¹⁷¹, European Industrial Technology Roadmap¹⁷² and the New Telco Cloud Thematic Roadmap¹⁷³, which highlight the urgent need for an upgraded infrastructure capable of supporting the next generation of cloud and edge applications.

The 6G-IA Vision Working Group has published in November 2024 the second version of the “European Vision for the 6G Network Ecosystem white paper”¹⁷⁴. This white paper presents a comprehensive vision for the development and standardisation of 6G networks in Europe, with a target commercial launch around 2030. The paper emphasises the importance of global collaboration while highlighting the specific contributions and priorities of the European 6G research and innovation landscape. Section 3 of the white paper defines eleven areas of technological enablers for the complete 6G vision, ranging from energy efficient technology to new hardware. Among these, four can be seen as more directly connected to the Telco Cloud-Edge as understood in this roadmap:

- **Edge-Cloud Continuum:** 6G networks will leverage a distributed computing architecture that spans from the edge to the cloud. Research and innovation in edge computing technologies, including edge orchestration, data management, and application optimisation, are essential for enabling low-latency and context-aware services in 6G.
- **Network Softwarisation and Disaggregation:** This involves the virtualisation and softwarisation of network functions, enabling greater flexibility, scalability, and agility in network deployment and management. Research and innovation in this area are crucial for building cloud-native 6G networks.
- **Network Intelligence:** AI and ML will play a pervasive role in 6G networks, enabling intelligent automation, optimisation, and management of network resources. Research and innovation in AI/ML algorithms, models, and frameworks for network optimisation and control are crucial for achieving the performance and efficiency goals of 6G.
- **Security:** As networks become more software-defined and distributed, ensuring security and trustworthiness becomes increasingly challenging. Research and innovation in

¹⁷¹ https://www.bmwk.de/Redaktion/EN/Downloads/II/ipcei-cis-value-chain-description.pdf?__blob=publicationFile&v=1

¹⁷² <https://digital-strategy.ec.europa.eu/en/news/european-alliance-industrial-data-edge-and-cloud-presents-its-first-deliverables>

¹⁷³ <https://digital-strategy.ec.europa.eu/en/news/new-telco-cloud-thematic-roadmap-european-alliance-industrial-data-edge-and-cloud>

¹⁷⁴ <https://doi.org/10.5281/zenodo.13708424>

security technologies, including zero-trust architectures, intent-based networking with AI assistance, and secure data sharing mechanisms, are essential for protecting 6G networks and services.

Moreover, the 6G-IA white paper also mentions specific technologies and approaches covered by this roadmap even if they are not yet to be considered at the core of Telco Cloud-Edge, such as Neuromorphic Computing (section 9.1) or Digital Twins of the network.

7.1 Open radio access network (Open RAN)

The radio access network (RAN) is mostly closed in terms of an “open development space” except for large OEM telecommunications companies. Unlike cloud native, where 3rd parties can develop microservices, advertise and sell/license, the RAN is a hardware/software monolith which is a hurdle to enter and offer services within and existing licences, limited standards access, and protective patents further limit the possibilities to enter the market. Antennas, base stations, MSCs, 3GPP and IETF protocol development, plus global testing are huge undertakings. One way to make the barrier to market entry lower is through the use of common off-the-shelf hardware, open software APIs accessible to anyone, as well as willing customers interested in easing entry into RAN & its development.

A way to achieve this is through Open RAN architectures decouple network hardware and software, enabling telecom operators to mix and match components from different vendors. This modularity allows for customized network solutions that can be tailored to specific operational needs and regional requirements. The main benefits of Open RAN include enhanced network flexibility, reduced vendor lock-in, and lower costs, facilitated by the use of standard, off-the-shelf hardware and open software interfaces. There is work being conducted on multiple levels of maturity and market readiness. Some of the key work that is likely to have high impact is done in the context of two alliances of players from the telco industry: the O-RAN Alliance¹⁷⁵ and the Telecom Infra Project (TIP)¹⁷⁶ which consolidate interested parties, identify missing parts and provide a base on which to build projects. There has a recent MoU signed by major European telecom operators that aims to deploy Open RAN technology across Europe to foster a competitive and innovative telecommunications ecosystem.¹⁷⁷ Open RAN as of 2024 fall into 2 categories: i) large players ‘promising’ to develop an open solution for the benefit of all, e.g., the Ericsson - AT&T \$14 billion deal,¹⁷⁸ or ii) smaller component-like development, a la Cloud native. Either way it remains to be seen whether either will deliver a truly open, competitive RAN.

In terms of government and regulatory support, various national and international regulatory bodies are beginning to support the Open RAN movement, advocating for policies that encourage open network architectures to enhance market competitiveness and innovation.

¹⁷⁵ <https://www.o-ran.org>

¹⁷⁶ <https://telecominfraproject.com>

¹⁷⁷ <https://www.politico.eu/wp-content/uploads/2021/01/20/POLITICO-Memorandum-of-Understanding-OPEN-RAN-big-four-operators-January-2021.pdf>

¹⁷⁸ <https://www.ericsson.com/en/news/2023/12/ericsson-and-att-in-major-future-network-of-the-future-deal>.

- **Global Coalition on Telecommunications (GCOT)** - This coalition includes countries like the UK, Australia, Canada, Japan, and the United States. In January 2025, they issued voluntary principles aimed at supporting an industry-led certification program for Open RAN products to help standardize and increase the deployment of Open RAN technologies.¹⁷⁹
- **Open RAN Policy Coalition** - The Open RAN Policy Coalition represents a group of companies formed to promote policies that will advance the adoption of open and interoperable solutions in the Radio Access Network (RAN) as a means to create innovation, spur competition and expand the supply chain for advanced wireless technologies including 5G. Coalition members believe that by standardizing or “opening” the protocols and interfaces between the various subcomponents (radios, hardware and software) in the RAN, we move to an environment where networks can be deployed with a more modular design without being dependent upon a single vendor.¹⁸⁰
- **National Policies** – Many countries have developed specific national strategies to promote the adoption of Open RAN. These strategies often include funding research and development, facilitating trials, and creating policies that encourage the use of open and interoperable network solutions.^{181 182}

To ensure competitiveness and leadership of Europe in this area, it is important to consider European relations to such initiatives as the above. Specifically, the recent statement from the GCOT includes many of the leading countries in telecommunications, but with the obvious omission of any EU Member States.

Challenges

As systems start using components from multiple vendors, there are multiple challenges in coordinating both their technical integration and assigning accountability between parties, such as when SLAs are unfulfilled¹⁸³, specifically:

- **Increased security threats** due to the increased system complexity and involvement of multiple vendors and 3rd party applications. It will be important to ensure that the system is protected, both the interfaces between different apps, and more importantly between 3rd party apps and the system as a whole.
- **Conflicts between RAN Intelligent Controller Apps:** Current AI/ML solutions typically target a single network layer and perform one intelligent network task. The development towards an AI native 5G and beyond network, we can expect a massive number of AI models co-existing and collaborating in a distributed manner, addressing different layers of the network and optimizing different KPIs. This complexity will increase risks of conflicts between different Intelligent Controller Apps, that need to be resolved and prevented.

¹⁷⁹ <https://globalregulatoryinsights.COTcom/art/global-coalition-on-telecommunications-open-ran-certification-principles/>

¹⁸⁰ <https://www.openranpolicy.org/about-us/>

¹⁸¹ <https://www.cullen-international.com/news/2023/09/Open-RAN--global-policy-and-industrial-perspectives.html>

¹⁸² <https://www.openranpolicy.org/governments-worldwide-embrace-open-ran-a-glimpse-of-recent-initiatives/>

¹⁸³ https://6g-ia.eu/wp-content/uploads/2024/05/6g-ia-open-sns_open-networks-status-and-future-development_ran-final.pdf

- **Cross layer optimisation:** As the number of RAN Intelligent Controller Apps increase, there will be many components triggering various actions to optimise RAN performance, such as triggering policies and modifying RAN configurations. To deal with this complexity, it will be necessary to introduce coordination mechanisms to ensure that the optimisations performed by the Apps do not degrade the overall system performance. This is especially interesting from the computing continuum perspective, which adds additional optimisation layers for data and compute orchestration.

Research & Innovation Priorities

- New security solutions for an open RAN ecosystem to develop cybersecurity solutions that can manage the multi-vendor security challenges of O-RAN and allow secure deployment of 3rd party apps in a secure manner.
- Cross-layer optimisation:
 - Joint optimisation of different RAN network layers and network tasks
 - Using RAN to optimize network-data-compute orchestration from device to cloud
- Develop and demonstrate reference architectures for how O-RAN could be used and integrated with Cloud-Edge orchestration solutions for different use cases.
 - O-RAN implementations, software and hardware co-design
- Simulation/validation environment to ensure that different RAN Intelligent Controller Apps play well together
- Identify parts outside the existing alliances that could be promoted as *topics for future EU calls* with OEM partnerships. Delve into the 6G and 7G generations that are still open before they are subsumed by the 4 major telecom players.
 - How to include developments related to seamless data connectivity described in section 7.2 in future O-RAN.

Potential Impact

Promote open and fair competition in the telecoms market. Lower entry cost for new operators resulting in a more competitive market. Make the possibility of European lead in telecoms whilst breaking the dominance of US hyperscalers in the Cloud market. Much more open source in telecommunications solutions in the future, à la SIP servers, Asterisk (Open PABXs). This does however not mean that we need more operators, on the contrary, we need fewer, but bigger, operators in Europe to compete with global cloud players.

7.2 Seamless data connectivity and predictive handover across different networks

Many industry sectors and use cases are increasingly adopting technologies from robotics and autonomous systems, such as drones, for various monitoring or delivery applications. In many applications they will need to cover large distances and/or operate in remote areas with limited dedicated communication infrastructure and bandwidth. Yet, they will need to maintain connectivity at all times with a human operator, and therefore an edge/cloud server,

possibly due to regulations. If the network experiences issues, or they move outside the range of the network, there is a risk of communication blackout when it must change to another network, such as between telco 4G/5G and satellite communication.

Seamless handover between networks could also be needed in case of network disruptions, caused either by human intervention or natural events. In such cases, it is important to quickly identify a new network channel and hand over communication from the primary communication channel with minimal disruptions.

Challenges

In terms of networking and communications, the European Union's connectivity infrastructure is still unprepared to meet the demands of a society and economy increasingly driven by data. For example, handover between common and future networks (4G, 5G, 6G, LoRaWAN, Wifi7, SatCom, StarLink) is not totally seamless, causing disruptions in ongoing continuous data flows when moving between them.

Research & Innovation Priorities

A set of measures that will enable a more seamless handover between networks in the Compute Continuum:

- Secure base accessibility to all networks.
- Predictive and fast hardware communication channel switch, accompanied by channel switch software protocol.
 - Sensing and QoS detection of all physical network channels.
 - AI prediction of near future channel QoS.
- New mechanisms for predicting and responding to network handovers and failures.
 - Buffer and recovery mechanism on the edge devices to bridge switch over phase.
 - APIs and support for application design that can handle network disruption with graceful performance degradation and quick recovery on reestablishment of connectivity.
 - *Develop hardware and software stack* that enables easy integration of SatCom connectivity, utilizing multiple services in the background (Inmarsat, Iridium, Starlink, etc.).

Potential Impact

We are seeing a shift from standalone connected devices to highly interconnected systems, where products like cars and smart home devices communicate and integrate to offer enhanced services. This transformation, driven by digitization and AI integration, depends heavily on consistent data connectivity, as many devices rely on backend systems for AI processing and seamless service delivery. Thus, many applications require robust data connectivity to an edge node or operations centre. Ensuring seamless handover between networks and predicting network failures is a key factor in enabling a resilient connected society where IoT/edge devices or humans on the move continuously become disconnected from the cloud continuum.

7.3 Generative telco cloud

The Generative Telco Cloud refers to the use of cloud computing and generative AI in mobile networks to provide a transformative approach to the way telcos operate.

The telecommunications industry worldwide and specifically in Europe is facing significant growth and revenue challenges. As traditional revenue streams become stale, telcos need to find a way to improve their operational efficiency and reduce costs. This requires increasing automation and flexibility in their operations. Furthermore, the quick development of new technologies and the demand for high-speed connectivity are driving the need for robust cloud and network infrastructure. Europe needs to accelerate its cloud and network developments to support the telco industry's capabilities.^{184,185,186,187}

Challenges

The integration of AI and cloud technologies in the telecommunications industry poses significant challenges. Developing and deploying solutions for automation and AI to improve operational efficiency and reduce costs is a major challenge. Additionally, integrating services across the ecosystem, designing networks that can seamlessly incorporate mobile operators, cloud services, and other stakeholders, and meeting rising customer expectations for personalized services and real-time communication are also significant challenges.

Furthermore, the complexity of modern networks, coupled with the need for scalability and flexibility, poses challenges such as managing dynamic network traffic, ensuring network security and integrity, and providing high-quality services to a growing number of connected devices. Cloud technologies also raise concerns about data security, privacy, and compliance with regulatory requirements. To overcome these challenges, telecommunications providers must leverage AI-powered analytics, predictive maintenance, and optimization strategies to prevent service disruptions and maintain network performance.

Research & Innovation Priorities

To address the challenges associated with the integration of AI and cloud technologies in the telecommunications industry, the following R&D priorities have been identified:

- **Data Quality and Integration:** Developing solutions to integrate and process high-quality data from various sources to support AI-powered decision-making. This includes for example: developing data architectures that enable the integration of heterogeneous data sources, creating data processing algorithms that enable data cleaning, transformation, and loading, implementing big data technologies to handle large amounts of data or researching machine learning techniques to improve data quality.

¹⁸⁴ *State of AI in Telecommunications: 2024 Trends*

¹⁸⁵ *GSMA Open Gateway - Open Gateway*

¹⁸⁶ *Huawei drives innovation with AI-read... - Mobile World Live*

¹⁸⁷ *CNEC Global Leap Program Debuts in Barcelona, Highlighting Seven Tech Leaps to Accelerate Intelligence-Huawei Cloud*

- **Network Complexity and Scalability:** Researching and developing new network architectures and protocols to support the scalability and complexity of modern networks. This includes for example: developing network architectures that enable scalability and flexibility, creating communication protocols that enable interoperability between devices and networks, researching traffic engineering techniques to optimize network performance or developing network monitoring and management tools to improve efficiency.
- **Enhance the Domain Knowledge:** Developing and implementing AI-powered solutions that can interpret and understand network behavior and performance. This includes developing machine learning algorithms that enable the interpretation of network data, creating simulation models that enable the prediction of network behavior and researching data analysis techniques to improve understanding of network behavior.
- **Explainability and Transparency**¹⁸⁸: Researching and developing AI-powered solutions that provide transparent and explainable decision-making processes. This includes developing machine learning algorithms that enable the interpretation of decisions, creating simulation models that enable the visualization of decisions or researching data analysis techniques to improve understanding of decisions.
- **Security and Privacy:** Developing and implementing robust security and privacy measures to protect data and networks from potential threats. Security protocols are required to enable the protection of data and networks, as well as encryption algorithms that enable the protection of data and threat detection techniques to improve network security.
- **Generative AI and Network Management:** Researching and developing generative AI solutions for network modeling, optimization, and anomaly detection. This includes creating simulation models that enable the prediction of network behavior as well as researching data analysis techniques to improve anomaly detection in networks.

Potential Impact

The impact of AI in telecommunications is far-ranging,¹⁸⁹ and includes:

- **Efficiency and Scalability:** Telcos will enhance their flexibility, cost-effectiveness, and the ability to rapidly deploy and manage services, ensuring efficient infrastructure utilization and capacity management¹⁹⁰.
- **Customer Experience Improvement:** Chatbots and virtual assistants driven by AI will respond to consumer enquiries, resolve problems, and customise communications. Proactive service offerings will result from predictive analytics' ability to foresee client needs.¹⁹¹

¹⁸⁸ draft-irtf-nmrg-ai-challenges-04 - Research Challenges in Coupling Artificial Intelligence and Network Management

¹⁸⁹ The Impact of AI In Telecommunications | European Business Magazine, <https://europeanbusinessmagazine.com/business/the-impact-of-ai-in-telecommunications/>

¹⁹⁰ "A Primer on Generative AI for Telecom: From Theory to Practice", <https://arxiv.org/abs/2408.09031>

¹⁹¹ Unleashing the Power of Edge-Cloud Generative AI in Mobile Networks: A Survey of AIGC Services, <https://ieeexplore.ieee.org/document/10398474>

- **Revenue Growth and Cost Reduction:** Monetize AI by offering data analytics services to other industries. Insights derived from telco data can drive business decisions across sectors. AI-driven content recommendations can boost revenue by increasing user engagement.
- **Predictive Maintenance and Network Optimisation:** Telcos will be able to enhance network reliability by identifying potential issues before they impact services as well as automate processes, optimize network resources, and enhance operational efficiency.
- **Predictive Protection Actions:** User's exposure to specific threats, leveraging the information of threat frameworks to facilitate the adoption of preventive actions. Behaviour based analytics through artificial intelligence techniques will assess if the behaviour of a group of users is akin to the behaviour of a known threat changing Network level security posture dynamically.

7.4 Resilient on-prem 5G/6G edge-cloud for Industry 4.0 and beyond

Many industries are in the process of connecting their equipment to their IT infrastructure to gather data to enable predictive maintenance and inform decision making processes, coordinate work across the plant, and for using unmanned vehicles and robots. The seamless delivery of advanced data services and AI-based applications is integral to Industry 4.0. Necessitating robust on-premises 5G/6G edge cloud infrastructure to ensure resilience and reliability amidst the challenges of integrating legacy automation systems with cutting-edge cloud technologies. In many industrial applications resilience and reliability are critical. Key initiatives in this area are the telco global API alliance¹⁹² and the open source 5G Core network developed by the Magma initiative¹⁹³.

Challenges

The automation systems in industry often rely on old computer systems, making it difficult to seamlessly integrate them with modern cloud technologies and keep the systems up to date. Most of the systems need to be on-prem very close to the production line and it is therefore not possible to use public telecom and cloud resources that might suffer from short disconnects, making it necessary to have private on-premises 5G/6G edge cloud deployments.

Research & Innovation Priorities

- **Reference designs** for different environments and use cases.
- **Portability** between public telco operator environments and private 5G/6G environments, compatibility with APIs, and using cloud-native technologies.
- **Seamless management** of IoT devices that do not use 5G/6G.

¹⁹² <https://camaraproject.org>

¹⁹³ <https://mamacore.org>

Potential Impact

Easier to seamlessly port applications and services between private on-prem environments and public cloud-edge. Help boost the deployment and use of 5G technologies and edge computing.

8 Digitalisation of Industry Sectors: Requirements from Next-Gen Applications

The computing continuum will enable both new use cases and applications across various sectors and transform existing ones. The specific requirements that these pose on the Cloud-Edge infrastructure, and how it is operated and managed, will differ between use cases.

This section explores some of these use cases and their requirements.

8.1 Automotive and the continuum

Vehicles are becoming ‘computers on wheels’ with technology assistance for the drivers. Currently the assistance is often safety related. Fully automated tech. finds its way down the tech chain: cruise control, lane departure warnings, automatic crash signalling (OnStar in the US, Apple devices) to name a few.

Technology extends beyond the driver; computer-based maintenance, GPS-enabled theft electronic Vehicle Identification Numbers. An (e)SIM is commonplace in trucks, transmitting maintenance, wearage information to OEMs on behalf of hauliers that demand high uptimes using on demand service.

The industry is also changing *how* software is developed. Moving towards Software Defined Vehicle (SDV) architectures where features are updated continuously and uploaded over the air. Connectivity using cloud services offer new possibilities. Motivated by Tesla and Jenkins-style development, the SDV in the automotive sector spans the continuum from the cloud, edge and devices. Smart cities through roadside units for safety, road monitoring, autonomous driving exemplify a rich environment for the continuum.

Roadside Units (RSUs), as intermediary storage and compute units, the automotive sector illustrates well the continuum. Vehicles store and upload data via RSUs an example of a working system is automated toll booths, transmitting and paying for the journey via IEEE 802.11p, WiFi with priority bands. Further developments will be cameras indicating VRUs, unforeseen icy roads or accident hot spots. RSUs demonstrate the use of *edge compute* as opposed to solely *edge devices*. Note, data sovereignty and privacy are key issues.

Challenges

Edge compute has traditionally been seen as a networking concept, primarily introduced to reduce latency. Groups of devices, vehicles in this case, sense data and need local processing, privacy-sensitive face & number plate removal. An achievable priority is a set of libraries to do the V2X, image processing, data anonymisation & secure sharing and storage.

Research & Innovation Priorities

- **Non-inter-operability between non-compliant standards.** An example is 3GPP and the IEEE. Parts of 5G for vehicular communication and a variant of IEEE 802.11p, named G5 where simultaneously chosen by different OEMs in the automotive sector in Europe.

- This caused Cooperative Intelligent Transport Systems (C-ITS) systems and the CCAM EU frameworks for vehicle-to-infrastructure communication to fracture or significant uncertainty in directions to select. The impact on R&D was large, with indecision around which tech base to use in C-ITS and new areas emerging, multi-frequency antennas to be started. **Priority, be decisive and clear in future policy directions.**

Potential Impact

The automotive industry is one of the largest in Europe, and the emergent Computing Continuum can support many new services in the mobility sector.

8.2 Augmented Reality/Virtual Reality technologies

Extended reality (XR) technologies, such as augmented reality (AR) and virtual reality (VR), are expected to transform everyday life fundamentally, and enable new use cases and applications in industry. Due to the high data throughput and low latencies required to deliver a seamless VR/AR/XR experience, these technologies are among the main use cases for 5G and edge computing technologies.

Challenges

The main challenges will be managing the high data throughput and the low latencies required.

Research & Innovation Priorities

- **Consistency in Virtual Environments:** Develop algorithms to ensure seamless and consistent user experiences across decentralized networks.
- **Latency Reduction:** Deploy and reinforce telecom and network infrastructure to minimize latency for real-time interaction within VR environments.
- **AI Integration:** Integrate AI to dynamically adapt and optimize VR experiences based on user interaction and environmental changes.
- **Interoperability Standards:** Establish standards to ensure interoperability among diverse VR platforms and decentralized computing resources.
- **Secure Data Exchange:** Create secure protocols for spatial data exchange in decentralized VR applications to protect user privacy and data integrity.
- **Operational Datasets:** Develop open operational datasets for training and evaluating AI models within VR scenarios.

Potential Impact

VR/AR/XR are enabling new use cases and applications, with the potential of bringing great value to industry and enterprises.

8.3 Cyberspace and physical space fusion

Originating from Japan, the concept of "Society 5.0" envisions a connected cyberspace where AI surpasses human capabilities, feeding optimal outcomes back into the physical realm.

Challenges

Ensuring data privacy and security on the Cloud is paramount as AI systems process vast amounts of data. High latency can hinder performance, particularly in critical applications such as healthcare or autonomous vehicles, where real-time processing is essential. Additionally, AI systems must efficiently scale with increasing data volumes and user demand to maintain optimal performance. The computational power required for AI can be costly, especially with the utilization of large neural networks, thus affecting overall expenses. Interoperability is crucial as AI systems often need to integrate with various other technologies and data sources, necessitating seamless compatibility. Moreover, regulatory compliance, particularly regarding data protection regulations like GDPR in Europe, is vital for AI systems operating on the Cloud, emphasizing the importance of cross-border data flows and specialised legal expertise. Lastly, addressing ethical considerations surrounding AI's integration into society, alongside ensuring data quality and standardization, remains imperative to harness its full potential while mitigating risks and limitations.

Research & Innovation Priorities

- **Data Privacy and Security:** There are risks associated with data breaches and unauthorized access, which necessitate robust encryption and security protocols.
- **Latency and Performance:** Cloud infrastructure must be optimized to minimize latency and provide the necessary computational power.
- **Scalability:** Cloud platforms need to provide flexible and scalable resources to accommodate the growth of AI applications without compromising performance.
- **Cost:** Organizations must manage the cost of Cloud resources effectively to make AI integration economically viable.
- **Interoperability:** Ensuring interoperability between different Cloud services and AI models requires standardized protocols and interfaces.
- **Regulatory Compliance:** AI Governance platforms: Automate the identification of regulatory changes and translation into enforceable policies, Risk management and lifecycle governance.
- **Technical Expertise:** Training and recruiting talent is necessary to bridge this gap and drive integration forward.
- **Ethical Considerations:** Ensure that AI systems are transparent, explainable, and aligned with human values.
- **Data Quality and Standardization:** Ensure that shared data is of high quality and standardized for interoperability.

Potential Impact

The convergence of cyberspace and physical space envisages a seamless integration, facilitating smart cities and environments where data exchange between sensors and

cyberspace optimizes resources and enhances quality of life. Society 5.0 embraces AI-driven analysis, wherein AI not only processes data but also offers feedback and solutions, augmenting decision-making. This feedback loop, empowered by AI, is anticipated to spawn new value across sectors, fostering economic growth, job creation, and societal well-being.

9 Disruptive impact initiatives

In addition to the previous sections and topics in this document, the NexusForum project has identified several topics of interest that could have a potentially disruptive impact on the cognitive computing continuum in the longer term. These are:

- Neuromorphic systems
- Space edge
- Integration of quantum computing infrastructure
- Quantum and classical computing fusion

These topics and potential synergies with the computing continuum are further described in the following sections.

9.1 Neuromorphic systems

Traditional digital computers are struggling to keep up with the demands of advanced technologies and applications, such as AI and Edge computing, as well as the optimisation of intricate systems. This is evident in simulations used in chemistry and pharmaceuticals to speed up the development of new medications and vaccines, in the process optimisation of complex production lines, as well as in the area of traffic and freight transport.

Neuromorphic computing, which aims to emulate the self-organizing and self-learning nature of the brain, offers a promising solution to this challenge. Despite its potential, neuromorphic hardware has not found its way into commercial AI data centres. One reason is due to insufficient opportunities for application-oriented testing of the hardware developments required for the highly complex computing technologies, as well as for a rapid implementation of the results in prototypes and small series. Another is the lack of standardized neuron models and common training techniques.

As an example, the EU-funded HYBRAIN¹⁹⁴ project aims to develop a new computing system that is inspired by the human brain. This system will be based on 'ultra-fast response' technologies, combining a number of highly innovative solutions based on integrating complementary technology platforms. Companies like SynSense and Innatera are designing neuromorphic hardware specifically for integration into consumer devices, for sensor-adjacent processing. They have deployed several audio inference use cases to Xylo Audio, and a bunch of vision inference use cases to Speck. Moreover, the first EU exascale supercomputer JUPITER will include a neuromorphic module.

Neuromorphic computing introduces a *new event-based paradigm* and new ways of developing and deploying AI applications. It is based on a different architecture than the commonly used von Neumann architecture, which means that existing software tools and solutions are generally not compatible with it. The event-based nature of neuromorphic computing systems

¹⁹⁴ <https://cordis.europa.eu/project/id/101046878>, <https://hybrain.eu/>

will have new requirements on the computing continuum, since the communication and computing patterns will differ from those of traditional systems.

Increased efforts in *hardware-software codesign* can accelerate the seamless integration of neuromorphic computing capabilities in the Compute Continuum. This will require adapting existing software stacks and architecture patterns commonly used today and creating novel interfaces between the neuromorphic and digital computing domains.

The integration of neuromorphic hardware into data centres and at the edge could lead to more sustainable and energy-efficient AI processing. This could revolutionize the processing of sensor data at the edge, opening new possibilities for applications in wearables, smart home, and IoT devices. The development and improvement of neuromorphic software can lead to more efficient and powerful AI systems. It can enhance the overall learning performance for specific tasks, moving away from hardware benefits to understanding the potential application benefits of neuromorphic computing. It can also lead to improvements in neuroscience as researchers start to recreate our grey matter in electronics.

With its ultra-efficient, event-driven processing capabilities, neuromorphic hardware enables unprecedented power-performance gains for always-on sensing use cases. This could lead to significant advancements in various fields, including chemistry, pharmaceuticals, production lines, traffic and freight transport, and more. Other communities, such as ESA, are also exploring neuromorphic-oriented calls. Event-driven processing, edge computing, and GDPR compliance are essential considerations for successful integration.

9.2 Space Edge

The European Space Agency (ESA) has launched several relevant initiatives on the topic of edge computing and AI in space. For example, one of ESA's Phi Labs is exploring Cognitive Cloud Computing in Space (3CS),¹⁹⁵ and another one focusing on edge learning in space is anticipated.¹⁹⁶ Their Phi-sat programme performs experiments to deploy on-board AI on satellites for various earth observation tasks, such as filtering out high-value, for example cloud-free, images before they are sent back to earth.^{197,198}

At the moment, it seems that both satellite integration and space edge in the computing continuum are firmly in scope for the SNS JU, but there are potential synergies and points of collaboration between the SNS projects, and the projects funded under CL4/DATA- or CL4/DIGITAL-EMERGING- calls. Beyond this, there are several other points that suggest a convergence and synergy:

- Edge Computing, with or without “Space” in front, is ubiquitous.
- Flexible service orchestration is also quite prominent in SNS JU.

¹⁹⁵ <https://philab.esa.int/esa-explores-cognitive-computing-in-space-with-fdl-breakthrough-experiments/>

¹⁹⁶ <https://philab.esa.int/esa-and-ai-sweden-sign-letter-of-intent-for-first-pilot-of-esa-%cf%86-labnet/>

¹⁹⁷ https://www.esa.int/Applications/Observing_the_Earth/Ph-sat

¹⁹⁸ https://www.esa.int/Applications/Observing_the_Earth/Phsat-2/New_satellite_demonstrates_the_power_of_AI_for_Earth_observation

- Virtualisation and Cloud Computing techniques (e.g., FaaS) overlap with Virtual Network Functions (VNFs).
- Data-driven AI is being considered, if not already adopted as a key enabler for 6G networks.

There are several ongoing projects and initiatives in this direction under the SNS-JU that could be particularly relevant to the computing continuum.

The SNS-JU RIA project 6G-NTN¹⁹⁹ (NTN project, using AI, Stream B, Communication Infrastructure Technologies and Devices) covers the complete integration of Terrestrial Network (TN) and Non-Terrestrial Network (NTN) in 6G (5G considers TN and NTN as two separate segments and does not try to cross-optimize between them). The NTN does not consider only satellites, but anything above the surface of the earth i.e. airplanes, drones, and similar airborne vehicles. There is also mention of space edge in some of its use cases:

- Public deliverable “D2.1 Use Cases Definition” (mention of space edge on page 13, page 15, page 16, page 61)²⁰⁰
- Public deliverable “D3.1: Report on 3D multi layered NTN architecture”, 1st version (description and assessment of the various components of the multi-layered NTN)²⁰¹

The 6G-NTN project already has space edge as one of its potential innovations, particularly relevant to the project use cases, UC3 Urban Air Mobility and UC4 Adaptation to PPDR or Temporary Events.

The projects answering to the newly closed SNS-JU call²⁰² HORIZON-JU-SNS-2024-STREAM-B-01-03 would preferably expand the work of projects like 6G-NTN. The call mentions for example: “Disaggregation, and virtualisation considering the ground and non-terrestrial segments are in scope. It should enable integrated space and ground edge computing and in-space traffic decision procedures allowing a ‘router in the space’”. There are also mentions of IoT-to-satellite connectivity.

9.3 Integration of quantum computing infrastructure

Quantum computing is emerging as a transformative technology with the potential to revolutionize various sectors, including cryptography, materials science, and complex system simulations. Globally, significant investments are being made to develop scalable quantum hardware and efficient quantum algorithms. Countries like the United States and China are

¹⁹⁹ <https://www.6g-ntn.eu/>

²⁰⁰ https://www.6g-ntn.eu/download/d2-1-use-case-report/?wpdmdl=2866&masterkey=gYhW0HmX3ZCCL6U6pyUC44YLR3-ifgCr20jzkviQOnpxrHiOtv-V61zVILhuYBtWaz7_ngTLvq1A79el2guA711VxQiqS541ESPnDd0XB4

²⁰¹ https://www.6g-ntn.eu/download/d3-1-report-on-3d-multi-layered-ntn-architecture-1st-version/?wpdmdl=2870&masterkey=jaeYz8R8ZyqzKPzsuAKYsXkZS_cjiWBcx1f4PEz2UhPW_Zs8ctqVUUu0Bcl4YwiB-jrxWWBNvmvS5HNAuj_0_wFrH-Zd0QL2aOlPcAAVF9A

²⁰² <https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/horizon-ju-sns-2024-stream-b-01-03>

leading the charge, aiming to achieve quantum supremacy and integrate quantum computing with existing HPC and cloud infrastructures.

In Europe, the focus is on building a robust quantum computing ecosystem that leverages existing strengths in HPC and fosters innovation across the continent. The European Union has launched several initiatives to support the development and integration of quantum computing technologies. These efforts aim to position Europe as a leader in the global quantum race, ensuring that European industries and researchers can fully exploit the capabilities of quantum computing. Europe is developing different types of quantum computers, including those based on superconducting qubits, trapped ions, and photonics, to explore various approaches and applications. Key initiatives include the establishment of European Quantum Excellence Centres, which aim to foster the development of an ecosystem of quantum programming facilities, application libraries, and a skilled workforce.

The current infrastructure for quantum computing (QC) is limited, making it difficult to fully leverage the capabilities of QC and integrate it with HPC and cloud services to enable efficient AI workloads. Quantum computing is a promising technology that could revolutionize various sectors. However, the lack of a robust infrastructure, including hardware and software, hinders its full potential.

The challenges include developing scalable quantum hardware, creating efficient quantum algorithms, and integrating QC with existing technologies like HPC and cloud services. Some relevant current initiatives include:

- HPCQC EU Project: High Performance Computer – Quantum Simulator hybrid.
- OpenSuperQ: A Quantum Computer for Europe.
- Quantum Flagship (and QUCATS): One of the most ambitious long-term research and innovation initiatives of the European Commission.
- Six EuroHPC sites selected to host first European quantum computers.²⁰³
- QuIC – Quantum Industry Consortium: QuIC is a pan-European non-profit association dedicated to the advancement of commercial quantum solutions and Europe's competitiveness in quantum technology on the global stage.
- QIA – Quantum Internet Alliance: QIA is a consortium of approx. 40 world-leading institutions working together to build a global Quantum Internet made in Europe.
- The European Quantum Communication Infrastructure (EuroQCI) Initiative.
- The QC4EO initiative from ESA exploring using HPC+QC workflows for a climate adaptation digital twin.²⁰⁴

The European Quantum Excellence Centres aim to foster the development of an ecosystem of quantum programming facilities, application libraries, and a skilled workforce.²⁰⁵ They will serve as a one-stop-shop for industry, academia, and the wider quantum technology user

²⁰³ https://eurohpc-ju.europa.eu/selection-six-sites-host-first-european-quantum-computers-2022-10-04_en.

²⁰⁴ <https://eo4society.esa.int/projects/qc4eo-study/>

²⁰⁵ https://eurohpc-ju.europa.eu/european-quantum-excellence-centres-qecs-applications-science-and-industry_en

community, accelerating the discovery of new quantum-oriented applications and fostering their knowledge and uptake.

The European Commission could fund projects focusing on addressing research challenges in quantum computing, the development of large-scale quantum computing testbeds, the creation of efficient Open Science quantum algorithms, and the integration of QC with HPC and cloud services. This could involve the use of the Modular Supercomputing architecture in the first EU exascale supercomputer JUPITER, which includes quantum and neuromorphic modules.

By funding these areas, the EC could significantly enhance Europe's computing capabilities, including quantum computing, leading to advancements in various sectors. This could foster innovation, create jobs, and strengthen Europe's position as a global leader in technology. It could also help address societal challenges, such as climate change and healthcare, by enabling more sophisticated data analysis and decision-making. It could also foster innovation, create jobs, and strengthen Europe's position as a global leader in technology.

9.4 Quantum and classical computing fusion

The fusion of quantum and classical computing represents a significant advancement in computational capabilities. Quantum computing uses qubits, which can process a richer set of possibilities compared to classical computing's binary bits. This makes quantum computing more efficient for certain types of problems, such as those involving exponential variables. The most promising algorithms are hybrid, combining quantum and classical approaches to leverage the strengths of both paradigms.

The primary challenge lies in the inherent limitations of classical computing, which uses binary bits and struggles with problems involving exponential variables. Quantum computing, on the other hand, uses qubits and can process a richer set of possibilities, making it more efficient for certain types of problems. However, quantum computers require specialised environments and operate at cryogenic temperatures, which makes them impractical for personal devices like laptops.

The solution to making quantum computing widely accessible is through cloud computing. Quantum calculations will be performed in data centres, and the results will be delivered through the cloud. This approach will allow the masses to benefit from quantum computing without the need for personal devices to handle the actual quantum processes. Hybrid algorithms, which integrate quantum and classical computing, are particularly promising. These algorithms can optimize the use of both computing paradigms, enhancing efficiency and performance. The fusion of Quantum Computing and Classical Computing in the cloud represents an exciting frontier in computing, promising to address problems currently beyond our reach and driving innovation and problem-solving in the future.

The convergence of quantum and classical computing in the cloud is expected to exponentially increase computational capabilities. This integration will enable software to determine the most efficient computing paradigm - classical, AI, or quantum - for different parts of a computation. Users will benefit from this combined power without being directly involved in the complex processes.

10 Conclusions

The European vision of a Cognitive Computing Continuum is an interconnected, federated, multi-provider environment where data and AI technologies seamlessly integrate across cloud, edge, and telco infrastructures. Creating a Cognitive Computing Continuum in Europe is vital for securing Europe's digital sovereignty, economic competitiveness, and technological leadership in an increasingly data-driven global economy. Achieving this will significantly reduce market fragmentation, mitigate dependency risks associated with non-European technology providers, and ensure compliance with Europe's regulatory and ethical frameworks.

This roadmap identifies several research & innovation priorities, complementing the work of major European initiatives and investments, towards realising a Cognitive Computing Continuum. It underscores the need for strengthening foundational capabilities in cybersecurity, interoperability, and open-source technologies, within Europe. Advancing interoperability and adopting open standards will be necessary to allow seamless integration across European providers and minimizing vendor lock-in.

The rapid growth of AI technologies and establishment of new data centers should consider the European energy supply and commitments to sustainability. Therefore, it is critical to invest in new technologies to advance energy efficiency across the computing continuum and adopt sustainable practices—such as carbon-aware computing strategies, advanced cooling solutions, and leveraging renewable energy—to position Europe as a global leader in energy-efficient environmentally responsible computing technologies.

Continued investments in establishing a leading European semiconductor industry, based on for example RISC-V, and focusing on developing next generation processor technologies, including optical computing, neuromorphic computing and quantum computing, will drive European leadership in next-generation computing technologies. Parallel investments should focus on driving market uptake and demand for these technologies, as well as creating software developer tools and middleware tools to deal with this increased platform heterogeneity.

The adoption of advanced digital technologies in specific industry sectors will be key to a competitive European industry. It is important to identify which ones are ready for adopting digital advances, and strike a balance between finding solutions that solve problems, and introducing very advanced technologies.

Europe needs to accelerate the deployment of AI technologies, intelligent orchestration in the Cognitive Computing Continuum, and next-generation telecommunications infrastructure to drive Europe's leadership in AI technologies and ensure readiness for emerging applications in the industrial digital transformation. Finding solutions for integrating existing EuroHPC resources and the coming AI Factories in the Cognitive Computing Continuum should be a key concern in the short to mid term.

Strategic collaboration between industry, academia, policymakers, and international partners, particularly South Korea and Japan, will enhance Europe's technological capabilities, innovation potential, and market reach. Europe needs a coordinated effort to stimulate

market demand, scale innovative SMEs, and provide clear regulatory frameworks that enable rapid technological adoption and secure investments.

By committing to this strategic direction, Europe and its strategic partners can realize this vision of a thriving Cognitive Computing Continuum—one that empowers citizens, businesses, and public institutions, advancing digital autonomy and leadership in the global digital landscape.



Consolidating Research and Policy along the Cognitive Computing Continuum



eucloudedgeiot.eu



Funded by the European Union

Project funded by



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs,
Education and Research EAER
State Secretariat for Education,
Research and Innovation SERI