European Parliament

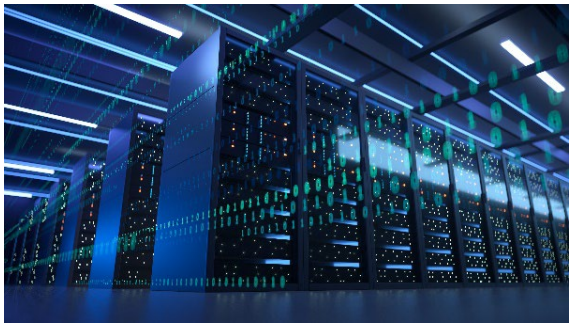
AI factories

SUMMARY

The EU has made significant progress towards trustworthy and ethical artificial intelligence that puts people first. In June 2024, the EU co-legislators signed the Artificial Intelligence Act (AI Act), the first horizontal legislation regulating AI systems. The focus is now shifting to bolstering the EU's position in AI innovation. European Commission President Ursula von der Leyen has promised to establish AI factories within the first 100 days of her term. AI factories will give companies and researchers access to the EU's supercomputers, tailored to AI needs.

AI factories bring together three essential components: supercomputers, data and human capital. The European High Performance Computing Joint Undertaking plays a pivotal role in this initiative, providing the necessary supercomputing infrastructure and covering half of the acquisition and operation costs of AI-optimised supercomputers as well as half of the cost of the services provided by AI factories. Seven consortia across the EU were selected to establish these factories, in Finland, Luxembourg, Sweden, Germany, Italy, Spain and Greece. Further proposals for AI factories are expected in February 2025.

AI factories contribute to the EU's strategic autonomy and support AI start-ups, industry and academia by enabling AI development and innovation in the EU while helping to reduce AI model training time and costs. They also facilitate access to data, help to develop the necessary skills and provide professional support services. However, challenges remain, such as making the process of applying for computing resources as smooth as possible, continuing to improve the energy efficiency of supercomputers and associated data centres, ensuring timely access to state-of-the-art AI chips, and advancing quickly with the establishment of AI factories. If addressed successfully, AI factories have the potential to enhance AI innovation in the EU. At the same time, it is essential to act on other fronts, such as improving the EU's cloud-computing infrastructure.



IN THIS BRIEFING

EPRS | European Parliamentary Research Service

Author: Maria Niestadt
Members' Research Service

PE 769.492 – February 2025

Introduction

Artificial intelligence (AI) has advanced remarkably in recent years, transforming industries and societies worldwide. Among the most transformative developments have been in the field of general-purpose AI (GPAI), capable of performing various tasks such as generating new content in response to a user prompt – a subset also referred to as generative AI. These developments would have not been possible without specialised high performance computing capacity, massive datasets, large investment and expertise; here the EU possesses both strengths and weaknesses.

The EU has invested in the European High Performance Computing Joint Undertaking (EuroHPC JU) network of supercomputers, now being adapted to train large-scale, general-purpose AI models and emerging AI applications. Deploying a supercomputing infrastructure at the EU level is critical for maintaining autonomy from foreign providers and encouraging European companies to develop general-purpose AI models. Specifically, small companies and start-ups often face barriers in accessing specific computing resources necessary for AI development.

Training general-purpose AI models requires extensive computation on specialised hardware such as graphics processing units (GPUs). Originally developed to process images in the gaming industry, these chips excel at processing multiple small tasks simultaneously, making them essential for general-purpose AI model training. GPUs enable acceleration of data-intensive tasks such as machine learning. However, GPUs are expensive, often in short supply and designed and fabricated by a few non-European companies. For example, United States-based Nvidia reportedly controlled more than 95 % of the GPU market in February 2024. Nvidia also provides a software platform called CUDA[1] that helps developers to speed up their applications by using the power of GPU accelerators.[2]

To overcome hardware and software limitations, the EU start-ups frequently rely on foreign cloud computing providers to train, run and deploy their generative AI solutions. For example, the French AI champion Mistral signed a deal with Microsoft in 2024 to access its supercomputers to train its large language models and to distribute its AI models on the Azure platform. In February 2025, Mistral released its own AI assistant, 'Le Chat'. However, access to foreign cloud computing resources is expensive, particularly for emerging AI start-ups. At the same time, competitive offers from European cloud providers are limited.

The EU's network of supercomputers proposes a cost-free alternative for European small and medium-sized enterprises (SMEs) and start-ups, although this access comes with conditions, such as the requirement to report on results. The EU's supercomputers have already supported AI innovation, such as the translation platform Unbabel, which used a supercomputer to develop[3] a large language model that supports 24 official EU languages.

The availability of large, varied and high-quality datasets is another bottleneck in the EU. While the EU's General Data Protection Regulation (GDPR) ensures protection of personal data, it can limit the availability and usability of large datasets required for training generative AI models. For instance, one study found that the GDPR reduced data storage by 26 % and data processing by 15 % among EU firms, compared to their counterparts in the US. The EU has introduced initiatives such as the European data strategy, as well as European Common Data Spaces and legislative measures such as the Data Act, which aims to unlock data generated by connected products and services.

In terms of talent, the EU has world-class universities and research institutions. Programmes such as Horizon Europe and initiatives such as the European AI Alliance support collaboration and innovation. However, global demand for AI experts often outpaces the EU's ability to attract and retain talent. Factors such as lower pay, long hiring processes and visa complications hinder EU competitiveness.

AI factories aim to address some of these challenges and leverage the EU's strengths in AI. In her political guidelines for 2024-2029, von der Leyen promised to give AI start-ups and industry access to tailored supercomputing capacity. This initiative aims to improve not only access to sufficient

computer power but also to data and expertise. The AI factories initiative will complement other EU action designed to ensure that AI remains human-centric, ethical and trustworthy.

While an important initiative, AI factories alone will not fully resolve all of the EU's challenges in this field. As the Draghi report highlights, the EU should also address the limited availability of cloud computing infrastructure in the EU. Some[4] argue that that cloud-based computing platforms offer more flexibility for diverse AI development needs. For this reason it is important to combine the strengths of supercomputers with the advantages of cloud computing. In addition, it essential to progress on the capital markets union to make it easier for investors in one EU country to invest in AI companies in another.

What are AI factories?

The European Commission describes AI factories as dynamic ecosystems formed around European supercomputers that bring together critical components that facilitate the development of cutting-edge generative AI models. These components are the following: AI-dedicated supercomputers (see box), associated data centres in proximity or connected via high-speed networks, and talent, including data specialists, researchers and SMEs. AI factories will also provide training. Supercomputing support services centres are also an essential part of AI factories, facilitating access to supercomputers and to dedicated supercomputer-friendly programming facilities and algorithmic support.

Supercomputers are high-performing computing (HPC) systems with very high computational power capable of performing complex and large-scale computational tasks. They can process and analyse large volumes of data at a rate that exceeds other computers by far. Supercomputers require a scaled-up software that organises, assigns, stores and processes data in a particular way. A typical supercomputer has many general-purpose central processing units (CPUs), similar to those found in personal computers. In supercomputers, these general-purpose CPUs are connected via ultrafast networks. Supercomputers adapted to AI development also contain graphic-processing units (GPUs).

Sources: European Commission and *The Economist*.

AI factories can also benefit from other EU initiatives, including AI Testing and Experimentation Facilities (TEFs) and European Digital Innovation Hubs. The TEFs are networks of physical and virtual facilities that allow technology providers to test and validate the latest AI solutions prior to real-world deployment. European Digital Innovation Hubs are one-stop shops that support businesses and public sector organisations in their digital transformation. They offer access to technical expertise and testing, financing advice, training and skills development.

The role of the EuroHPC JU in AI factories

The EuroHPC JU is instrumental in establishing AI factories. Founded in 2018,[5] this institutional partnership develops supercomputing infrastructure and advances high-performance computing in the EU. The EuroHPC JU was established to address the EU's lag in supercomputer capabilities. Over time, its mandate has been extended to include quantum computing (since 2021) and AI factories (since 2024).[6] Today, it is an autonomous legal and funding entity, located in Luxembourg.

The EuroHPC provides the necessary supercomputing and data infrastructure for the establishment of AI factories. Its mandate allows it to acquire and operate AI-optimised supercomputers or upgrade existing EuroHPC supercomputers to AI capabilities.

The EuroHPC currently hosts nine interconnected supercomputers across the EU: Lumi in Finland, Leonardo in Italy, MareNostrum 5 in Spain, MeluXina in Luxembourg, Karolina in Czechia, Discoverer in Bulgaria, Vega in Slovenia, Deucalion in Portugal and Jupiter in Germany. Several of these supercomputers rank among the world's most powerful. The EuroHPC supercomputers are available to a large range of users, such as start-ups, SMEs, researchers and public authorities. AI-optimised

supercomputers[7] can be used for training large-scale, general-purpose AI models and emerging AI applications.

The EuroHPC JU is responsible for launching calls for expression for selecting hosting entities that host and operate AI-optimised supercomputers and AI factories. It evaluates applications, selects the hosting entities and signs the hosting agreements.

The EuroHPC JU acts as a single EU-level point of contact, helping users to the right service centre. Additionally, the EuroHPC supports skills development, promotes the take-up and use of European technology solutions, and supports collaboration between various AI factories and other AI initiatives.

Progress towards AI factories

In 2024, the Council amended the EuroHPC JU Regulation to enable the joint undertaking to establish and manage AI factories. This amendment was part of the EU's broader AI innovation package. The amended EuroHPC JU regulation allows the joint undertaking to buy and operate AI-optimised supercomputers or upgrade existing supercomputers with AI capabilities. The amended regulation widens the circle of supercomputers users to a large number of public and private users, including start-ups and SMEs with a specific focus on AI needs.

In September 2024, the EuroHPC launched two calls for expression of interest to select entities to host and operate AI-optimised supercomputers, AI experimental platforms and AI factories. The calls are open until 31 December 2025, with the first deadline on 4 November 2024 and subsequent deadlines every three months, as long as the funds through Horizon Europe and the Digital Europe programme are available.

Applicants had to outline a comprehensive AI factory implementation plan, including user support, skills plan and networking with other European and national initiatives or other AI factories. The calls for expression of interest envisage three options for establishing an AI factory; development around:

1. a newly acquired supercomputer;
2. an existing supercomputer that needs to be upgraded to AI capabilities;
3. a supercomputer that already has sufficient computing resources for training large-scale, general-purpose AI models and emerging AI applications.

Hosting entities could also choose to develop AI-experimental platforms to develop and test new ideas and innovations. The EU covers up to half of the development and operating costs of such a platform.

By November 2024, the Commission announced that the EuroHPC had received seven proposals for AI factories. The proposals were submitted by the hosting sites in Finland,[8] Luxembourg, Sweden, Germany, Italy,[9] Spain[10] and Greece. In addition, Cyprus and Slovenia indicated by letter their interest in either joining or creating an AI factory at a later stage.

By December 2024, the EuroHPC JU had given a green light to all seven proposals. Five hosting sites in Finland, Germany, Italy, Luxembourg and Sweden chose to develop an AI factory around a new AI-optimised supercomputer. In some cases, the selected consortium will first upgrade an existing system to meet the needs of AI-related research and innovation. The AI factory in Spain will update an existing supercomputer and Greece will establish its AI factory around a supercomputer that is currently under deployment. The AI factories in Finland and Spain also chose to develop an AI-experimental platform. Further proposals are expected in February 2025.

Financing AI factories

The financial resources for the establishment of AI factories come from both EU programmes and individual EU Member States. The EuroHPC covers up to half of the acquisition and operation costs of AI-optimised supercomputers, up to half of the costs of setting up and operating the AI factories

and half the costs of providing AI factory services. The hosting state or the hosting consortium covers the remaining costs. The EuroHPC and the hosting state (or consortium) also share access (compute) time to the supercomputer.

For the acquisition of new or upgraded AI supercomputers, the EU's total financial contribution was initially estimated at €400 million (Digital Europe funds) in 2024, and up to €800 million in total. Horizon Europe dedicates a further maximum €180 million to setting up AI factories and developing AI-experimental platforms. The EU also set an estimated maximum €200 million for the EU's contribution per new or upgraded supercomputer, as well as €15 million for three years for establishing and running each AI factory. The maximum amount may be adapted, depending on the number of applications received and available EU funds. At the AI Action Summit in Paris on 11 February 2025, President von der Leyen announced the EU will invest €10 billion in AI factories.

In comparison, the United States announced Stargate, a private sector project to invest US$500 billion in AI infrastructure. According to declarations made by the companies behind the Stargate project, the project has already started with the construction of a data centre in Texas. Unlike the EU's AI factories initiative, the Stargate project does not include public funds, but the US government is likely supporting it through land-use permits, and by ensuring access to cheap energy and water for the data centres.

Benefits and challenges

AI factories contribute to EU's strategic autonomy and facilitate an AI-innovation ecosystem. They facilitate access to the necessary computing infrastructure. According to the Secretary-General of the European Digital SME Alliance, this is a critical element in helping SMEs to build AI models and solutions. AI factories reduce AI-model training time and costs, as well as facilitating access to data. In addition, AI factories support the development of trustworthy AI, as access to EU supercomputers is only granted for the development of ethical and responsible AI models and solutions. They also improve AI skills by providing training that equips all parties involved with the skills needed to use the EuroHPC supercomputers for AI-model training and application development.

However, there are also challenges on the horizon. Start-ups are sometimes reluctant to apply for access to the EU's supercomputers due to long and complex application processes. According to a 2024 report, 12 % of start-ups found the bureaucratic effort to access the EuroHPC supercomputers too high, while 11 % were uncertain about the specific benefits and considered waiting times to gain access to the EU's supercomputers were too long, and 8 % were not sufficiently familiar with the EU's supercomputers. Some argue that some ambiguity exists around the concept of 'trustworthy AI'. Fear of non-compliance with the requirement to develop ethical and responsible AI might discourage some start-ups from participating in AI factories. The European Parliament has also highlighted the need to communicate the opportunities offered by AI factories to start-ups, SMEs and researchers more widely.

To address the bureaucracy issue, the EU has taken steps to facilitate and speed up the process. The EuroHPC JU website explains the application process in detail and national competence centres, and EuroHPC services offer additional guidance. A single point of contact at the EuroHPC JU helps start-ups and other interested users to contact the specific service centre, and every service centre has a one-stop shop for start-ups to facilitate access. In addition, the EuroHPC is currently updating the access policy to meet the demands of users with AI-specific needs.

It is equally important to ensure that energy-hungry supercomputers and associated data centres are energy efficient. Data centres, for example, account for about 1-1.5 % of global electricity use. The European Parliament's April 2024 resolution calls on supercomputer hosting entities to make concrete plans on the energy efficiency and environmental sustainability of supercomputers. In Parliament's view, those 'plans should ensure that the supercomputer has access to a secure and stable grid connection and electricity supply, preferably through clean affordable energy'. Energy

efficiency is crucial for economic as well as environmental reasons, as it can help to reduce the costs of powering supercomputers and data centres.

The EU's supercomputers have already taken a number of energy efficiency measures. In some cases, the waste heat from supercomputer water-cooling systems is used for local area heating. The supercomputer in Finland operates entirely on carbon-free hydroelectric renewable energy; Spain's is powered entirely by renewable energy; and Luxembourg's runs on green energy from a cogeneration plant powered by waste wood. Finland's supercomputer also benefits from its location in a colder climate, which helps to reduce cooling costs. However, it is also important that supercomputer users are aware of the sustainability of their operations.

Some analysts also see the availability of sufficient private funding and AI talent as potential challenges for AI factories. Establishing AI factories requires staff profiles such as data scientists, data engineers and machine learning engineers. Attracting and retaining this talent can be difficult.

Another potential challenge is continued access to chips (GPUs). Just before leaving office, US President Joe Biden issued new restrictions on the export of US-developed GPUs. These restrictions may affect some AI factories in countries that are not exempt from these restrictions, such as Luxembourg and Greece. At the same time, Chinese start-up DeepSeek recently argued it is possible to train and run industry-leading AI models at a fraction of the cost, using fewer advanced AI chips and less computation time than was necessary for models developed by Open AI, for example. This potentially puts the need for vast resources to build sophisticated AI models in question.

Conclusion

AI factories have the potential to enhance AI innovation in Europe, but it is important to address the challenges. Ensuring smooth access to supercomputers and GPUs, continuing to improve supercomputer and data centre energy efficiency and speeding up the establishment of AI factories are all crucial elements in making an AI factories initiative a success. There is a sense of urgency, as generative AI is evolving rapidly. To catch up with other regions, the EU needs to progress quickly.

In addition to establishing AI factories, the EU should also act on other fronts to remain competitive in AI, such as advancing on the capital markets union and addressing the limited availability of cloud-computing infrastructure in the EU. By taking a comprehensive approach that tackles multiple challenges, the EU has a better chance of becoming competitive in the global AI landscape.

MAIN REFERENCES

European Commission, Communication on boosting startups and innovation in trustworthy artificial intelligence, January 2024.

ENDNOTES

1 CUDA: Compute Unified Device Architecture.

2 GPU acceleration is a computing technique that uses the power of GPUs to accelerate the performance of applications substantially.

3 Together with other partners.

4 For example, the Lighthouse Europe consultancy.

5 The decision to establish the EuroHPC JU was made in Council Regulation 2018/1488.

6 Council Regulation (EU) 2021/1173 expanded the EuroHPC JU mandate to quantum computing and Council Regulation (EU) 2024/1732 to developing and operating AI factories.

7 According to Council Regulation 2024/173, an AI-optimised supercomputer is 'a supercomputer that is designed primarily for training large scale, general-purpose artificial intelligence ("AI") models and emerging AI applications'.

8 The proposal was presented together with Czechia, Denmark, Estonia, Norway and Poland.

9 The proposal was presented together with Austria and Slovenia.

10 The proposal was presented together with Portugal, Romania and Turkey.

DISCLAIMER AND COPYRIGHT

eprs@ep.europa.eu (contact)

https://eprs.in.ep.europa.eu (intranet)

www.europarl.europa.eu/thinktank (internet)

http://epthinktank.eu (blog)