POLICY BRIEF

Making Public Compute Work for Applied AI Startups

A Deep Dive Into Poland's AI Factory Ecosystem

Jan Króliński, Dr. Nicole Lemke

July 01, 2026

Tech analysis and policy ideas for Europe

interface

Table of Contents

Executive Summary

Globally, AI compute infrastructure has become a focal point of debates over technology sovereignty. The European debate surrounding public investments in AI infrastructure remains contentious, with many critics questioning whether sufficient private-sector demand exists to justify the scale of commitment. In this policy brief, we argue that the relevant question is not whether to build public compute, but for whom, in what form, and under what conditions it can boost European AI development.

Taking Poland as a case study – the fourth-largest national recipient of EU AI factory funding[1] and an emerging innovator according to the European Innovation Scoreboard (EIS) – we examine empirically whether public AI compute can stimulate commercial activity in a market that is not an obvious destination for frontier AI innovation. Based on ten semi-structured interviews with one AI factory host, four ecosystem enablers (industry associations and research bodies), and five AI startups operating in markets that are strategic for the EU, our findings show that:

- Demand for public compute is real but sector-specific. Applied AI startups in regulated and strategic sectors such as healthcare and life sciences, and space and defence have genuine compute needs that public infrastructure can address. Generalising from frontier-LLM developers to the entire universe of AI firms misrepresents this demand.
- Startups have concrete requirements from public infrastructure. No-cost access to high-performance computing (HPC) is welcomed, provided it is unbureaucratic, time-sensitive, protects data and IP, and enables easy switching between AI factories and commercial cloud providers. Localised human support and access to data are equally valued.
- Remaining challenges are real but not insurmountable. Securing national investment portions, bridging academic and commercial operating cultures, codifying sovereignty guarantees, and enabling portability and interoperability are the key obstacles. Policymakers and AI factory operators who address these directly are well placed to deliver on the investment's promise.
- Gigafactories are seen as a capacity upgrade, not a qualitative shift. Current applied AI startup priorities do not centre on gigafactory access; the operational and technical findings of this brief extend to gigafactories under broadly the same conditions. The more foundational challenges, including energy constraints, sit at the system level.

The central implication for policymakers is one of design: public compute infrastructure that is tailored to the requirements of applied AI startups, rather than optimised for frontier model training alone, is well placed to translate public investment into durable value for European AI innovation.

---

1 Tomasz Batko, 'W trzyetapowym konkursie Komisji Europejskiej (KE) zdecydowano o utworzeniu 19 Fabryk AI w 16 krajach Unii Europejskiej', LinkedIn, 2026, accessed May 14, 2026, https://www.linkedin.com/feed/update/urn:li:activity:7416928301056819200/.

Introduction

AI compute infrastructure has become a central point in global policy debates surrounding technology sovereignty. In the European Union (EU), the global competition over AI compute has led to the selection of nineteen AI factories and the announcement of several AI gigafactories. Moreover, the buildout of European compute infrastructure is one of the central goals of the EU's recently adopted tech sovereignty package.

At the same time, the debate surrounding public investment in AI compute infrastructure is highly contentious. The European AI compute debate currently sits between two positions. The first, which underpinned the original AI factory and gigafactory initiatives, treats public compute capacity as a precondition for European AI competitiveness as well as technology sovereignty, assuming that commercial demand will follow supply.[2] The second questions whether private-sector demand for European public compute infrastructure exists in sufficient volume to justify the scale of investment committed.[3] [4] Critics of public investment in AI compute often generalise from the trajectory of a small number of frontier-LLM developers to the entire universe of AI firms, and conclude that since few such firms are emerging at scale in Europe, public compute investment is misdirected.

Yet, firms developing frontier AI – such as France's Mistral.AI, which has recently gathered $830 million in debt for its own AI data centre build-up – are not the only ones in need of AI compute. AI adoption across a broad range of economic markets equally requires compute resources. In this policy brief, we argue that public compute infrastructure investment is best understood as a matching problem between different types of AI firms, different stages of AI deployment, and different infrastructure designs. The relevant question then is not whether to build public compute infrastructure, but for whom, in what form, and under what conditions it can boost European AI development. Building on previous interface research,[5] we seek to provide preliminary insights on how and under which conditions AI factories can contribute to commercial AI innovation in Europe.

Our argument does not focus on AI firms building frontier LLMs: these firms require

2 European Commission, 'EU launches InvestAI initiative to mobilise €200 billion of investment in artificial intelligence', *Shaping Europe's digital future*, February 11, 2025, accessed May 14, 2026, https://digital-strategy.ec.europa.eu/en/news/eu-launches-investai-initiative-mobilise-eu200-billion-investment-artificial-intelligence.

3 Pieter Haeck, 'EU accused of wasting €20B on AI computing dreams', *Politico*, May 4, 2026, accessed May 14, 2026, https://www.politico.eu/article/eu-accused-wasting-20-billion-euro-ai-computing-dreams/.

4 Julia Christina Hess and Felix Sieker, *Built for Purpose? Demand-Led Scenarios for Europe's AI Gigafactories*, Policy Brief, Interface and Bertelsmann Stiftung, October 22, 2025, accessed May 23, 2026, https://www.interface-eu.org/publications/ai-gigafactories.

5 Nicole Lemke and Catherine Schneider, *The European Union's AI Factories: Lessons for Public Investment in AI Infrastructure in Europe*, Policy Brief, Interface, October 30, 2025, 14, accessed May 15, 2026, https://www.interface-eu.org/publications/ai-factories.

extremely large-scale and highly specialised compute infrastructure, which is often privatised as the example of Mistral.AI shows. The AI startups interviewed for this policy brief are not building frontier LLMs; they are building applied AI products in highly regulated markets such as healthcare and life sciences and defence and space, both of which are considered of high strategic importance according to the EU's Apply AI strategy. Their demand for public compute infrastructure is real, sector-specific, and only partially aligned with what is currently being offered.

To examine the conditions under which AI startups and AI factories operate empirically, we take Poland as a case study. Poland is considered an emerging innovator according to the European Innovation Scoreboard (EIS).[6] Technological development is often path dependent, clustering in innovation ecosystems with a strong history of innovation in adjacent technologies.[7] Although Polish AI startups have recently shown genuine dynamism,[8] [9] its status as an emerging innovator makes it a particularly instructive case for testing whether public AI compute investment can generate genuine additionality, i.e., whether the investment creates new, additional value, rather than merely funding projects that would have happened regardless.

This is crucial since AI factories are meant to be more than providers of compute infrastructure. Rather than mainly serving existing demand, they are also designed to act as capability builders and knowledge brokers: supporting firms with expertise, connecting them with funders and other organisations, and thereby potentially stimulating new demand for the compute infrastructure itself. Poland is an ideal setting to explore how AI factories can fulfill these tasks in an environment where such additional support may be particularly needed. It is thus a particularly well suited test case for the European Commission's ambitions of AI factories boosting "dynamic ecosystems" for AI in Europe.

Case Selection and Methods

European public compute investment is currently structured around three pillars: The AI factories, representing public compute infrastructure, AI gigafactories, which are likely going to be developed in public-private partnerships, and the newly adopted tech sovereignty package, which aims to boost the buildout of European compute

6 European Commission, 'European Innovation Scoreboard 2025: Country profile Poland' (Brussels: European Commission, July 2025).

7 Flávio L. Pinheiro et al., 'The dark side of the geography of innovation: relatedness, complexity and regional inequality in Europe', *Regional Studies* (2022), https://www.tandfonline.com/doi/full/10.1080/00343404.2022.2106362.

8 Łukasz Badowski, 'Polskie startupy zarabiają miliony na AI. Technologia made in Poland podbija świat', *INNPoland*, March 12, 2026, accessed May 20, 2026, https://innpoland.pl/222118,polskie-startupy-zarabiaja-miliony-na-ai-technologia-made-in-poland-podbija-swiat.

9 PFR Ventures and Inovo.vc, *Polish VC Market Outlook Q1 2026*, PFR Ventures; Inovo.vc, 2026.

infrastructure and specifies sovereignty requirements for such infrastructure.

The nineteen AI factories selected so far are conceived as dynamic ecosystems built around AI-optimised supercomputers, providing computing resources and support services to European industry and to scientific users developing large AI models.[10] AI gigafactories, by contrast, will be designed to concentrate far larger compute resources[11] and are still expected to be selected through a European tender, although the timetable has been pushed back repeatedly and selection is now anticipated no earlier than Q4 2026.[12] AI factories and gigafactories also differ in purpose. AI factories are meant to build a broad ecosystem: their supercomputers are equipped to develop 'mainly middle-range AI models'[13] and to serve a wide base of researchers, SMEs, start-ups and the public sector. AI gigafactories target the next generation of frontier models; they are expected to require at least three to four times the most advanced AI processors currently available in the most powerful AI factories,[14] and therefore address a narrower set of users capable of designing and training very large-scale models.[15]

The programme was substantively changed in December 2025, when the Commission moved the procurement to the EuroHPC JU (European High Performance Computing Joint Undertaking – the EU body that jointly funds and coordinates Europe's supercomputing infrastructure with participating states, and which develops and operates the AI factories)[16] level and removed the option of designating a preferred consortium. Meanwhile, Latvia and Estonia withdrew from the Baltic AI gigafactory bid, leaving Poland, Lithuania and Czechia to continue under the revised rules.[17]

Additionally, the Commission hinted towards recalibrating the project from a training-centric framing toward explicitly including inference.[18] This shift is now reflected in law: the revised EuroHPC Regulation, adopted in January 2026, defines an AI gigafactory as a facility with sufficient capacity to handle "the complete lifecycle, from development to large-scale inference, of very large AI models and applications".[19] Reports on the programme's scale have, however, been contradictory: some suggested

10 European High Performance Computing Joint Undertaking, 'AI Factories', The European High Performance Computing Joint Undertaking (EuroHPC JU), accessed May 12, 2026, https://www.eurohpc-ju.europa.eu/ai-factories_en.

11 Hess and Sieker, *Built for Purpose?*

12 Nikola Bochyńska, 'KE zmienia terminy, Polska czeka. Niepewność wokół gigafabryki AI', *WNP.pl*, May 14, 2026, accessed May 15, 2026, https://www.wnp.pl/tech/ke-zmienia-terminy-polska-czeka-niepewnosc-wokol-gigafabryki-ai,1061717.html.

13 Council of the European Union, 'Council Regulation (EU) 2026/150 of 16 January 2026 amending Regulation (EU) 2021/1173 on establishing the European High Performance Computing Joint Undertaking', *Official Journal of the European Union* L, 2026/150 (January 19, 2026), recital 4, http://data.europa.eu/eli/reg/2026/150/oj.

14 Council of the European Union, 'Council Regulation (EU) 2026/150', recital 6.

15 Lemke and Schneider, *The European Union's AI Factories*, 14.

16 Council of the European Union, 'Council Regulation (EU) 2026/150', art. 1(2)(b)(ii).

17 Elżbieta Rutkowska, 'Polska może stać się hubem AI dla całego kontynentu. Wszystko zależy od tej decyzji', *WNP.pl*, May 7, 2026, accessed May 23, 2026, https://www.wnp.pl/tech/polska-moze-stac-sie-hubem-ai-dla-calego-kontynentu-wszystko-zalezy-od-tej-decyzji,1057662.html.

18 Maximilian Henning, 'EU recalibrates the case for AI gigafactories', *Euractiv*, April 7, 2026, accessed May 23, 2026, https://www.euractiv.com/news/eu-recalibrates-the-case-for-ai-gigafactories/.

19 Council of the European Union, 'Council Regulation (EU) 2026/150', art. 1(1)(a).

funding might ultimately cover only two sites rather than the five initially envisaged,[20] while more recent reporting indicates the Commission is now considering a more decentralised configuration of seven sites of varying size, four "standard" gigafactories of around 75,000 GPUs and three large-scale sites of around 100,000 GPUs, brought online in phases.[21] The shifting figures underline how unsettled the gigafactory design remains.

In addition to the AI factories and gigafactories, the Commission's recently presented proposal for the Cloud and AI Development Act introduces "Cloud and AI Leadership Initiatives", which are equally supposed to boost the buildout of large-scale digital infrastructure as well as bundle other offers helping to boost AI development, including frontier AI as well as AI in strategic sectors. Moreover, the Cloud and AI Development Act mentions the restructuring of European Digital Innovation Hubs (EDIHs) into experience and acceleration centres for AI. The concrete structure, offers, and governance of these initiatives and how they are going to be integrated with AI factories and gigafactories is an open but crucial question for the success of the EU's broader plans for supporting the construction of European compute capacity as well as AI development in Europe.

In this policy brief, we predominantly focuses on the nineteen AI factories that have already been selected all over Europe, in a diverse range of innovation ecosystems. Compared to the other pillars of the EU's strategy for boosting European compute infrastructure buildout, more information is available about AI factories, even though not all of them are operational yet. The conditions under which AI factories translate public compute investment into a boost for commercial AI innovation may shape the European AI landscape over the next several years and hold important learnings for AI gigafactories and cloud and AI leadership initiatives. We argue that AI factories must consider the needs of their innovation ecosystems and tailor their offer accordingly if they are to maximise their impact. This builds on and extends recent interface work,[22] which argued that AI factories' impact on commercial AI innovation will depend on how well their consortia, sectoral focus and ecosystem embedding align with the private actors they are meant to serve.

Complementary research[23] has emphasised that the geographical placement of factories overlaps only loosely with Europe's leading AI hubs of excellence, raising the question of whether infrastructure built outside already successful AI innovation ecosystems can

20 Luca Bertuzzi and Lewis Crofts, 'EU's AI gigafactories project plagued by tech and budgetary bottlenecks', *MLex*, April 13, 2026, accessed May 23, 2026, https://www.mlex.com/mlex/articles/2463944/eu-s-ai-gigafactories-project-plagued-by-tech-and-budgetary-bottlenecks.

21 Larissa Kögl, 'Aus fünf mach sieben: EU denkt KI-Gigafactories neu', *PoliticoPro*, May 28, 2026, accessed June 8, 2026, https://pro.politico.eu/news/aus-fuenf-mach-sieben-eu-denkt-ki-gigafactories-neu.

22 Lemke and Schneider, *The European Union's AI Factories*.

23 Nicoleta Kyosovska and Andrea Renda, *EU plans for AI (giga)factories: sanctuaries of innovation, or cathedrals in the desert?*, CEPS In-Depth Analysis, Centre for European Policy Studies, November 2025, accessed May 12, 2026, https://www.ceps.eu/ceps-publications/eu-plans-for-ai-gigafactories-sanctuaries-of-innovation-or-cathedrals-in-the-desert/.

generate the dynamic effects the Commission envisages. In our view, AI factories may serve different purposes depending on the type of innovation ecosystem they are embedded in: in strong existing AI innovation ecosystems, they may, for example, mainly serve existing demand for compute as well as strengthening existing AI capabilities. In emerging AI innovation ecosystems, they may instead be more relevant in building up capabilities and thus stimulating demand. Implemented correctly, the AI factories' diversity could thus develop into a genuine strength of the initiative.

We select Poland as our case study because the country is considered an emerging innovator by the EIS. Innovative activity is often path dependent, making Poland an interesting test case for the impact public compute investment can have on local innovation ecosystems precisely because of the more challenging conditions from which it is starting out compared to many other European innovation ecosystems. This allows us to explore whether – and how – AI factories can generate genuine additionality, new economic and innovative activity that would not have happened without them. The Polish AI ecosystem is relatively dynamic, suggesting that the necessary preconditions for such developments are in place.[24] [25] Since we are particularly interested in how public compute infrastructure can support AI startup activity in Europe – an explicit goal of the AI factories initiative – the present policy brief draws on ten semi-structured interviews with key actors across the Polish AI innovation ecosystem: one AI factory host organisation, five applied AI startups, and four organisations acting as ecosystem enablers.

While not fully representative, our interviews thus cover some of the most relevant organisations in the Polish AI innovation ecosystem. The AI factory represents one of two AI factories expected to come online in Poland: PIAST (PCSS, Poznań),[26] and GAIA (Cyfronet, Kraków).[27] Following the AI factories typology developed in previous interface work,[28] PIAST is a research-led regional consortium that functions as a capability builder, located in an emerging innovative region (Wielkopolska) (RIS - Regional Innovation Scoreboard 2025). Wielkopolska with its capital city of Poznań, where PIAST is located, is one of Poland's premier economic regions with a strong manufacturing focus. Foreign companies investing in the region are predominantly active in the automotive, food, transport and logistics, electronics, pharmaceuticals and beauty, and aerospace sectors, as well as business process outsourcing and shared services centres (BPO/SSC). Poznań itself is an important academic hub and the region is home to several incubators and industry and technology parks to support the creation of new

24 PFR Ventures and Inovo.vc, *Polish VC Market Outlook*.

25 Badowski, 'Polskie startupy zarabiają miliony na AI'.

26 Poznańskie Centrum Superkomputerowo-Sieciowe, 'PIAST-AI', accessed May 14, 2026, https://www.pcss.pl/piast-ai/.

27 Academic Computer Centre Cyfronet AGH, 'Gaia AI Factory', accessed May 14, 2026, https://gaia.plgrid.pl/.

28 Lemke and Schneider, *The European Union's AI Factories*.

companies.[29]

By the same criteria, GAIA is a research-led national consortium that combines capability building with sector-focused incubation in a moderate innovation region (Małopolska).[30] Małopolska with its capital Krakow is the only Polish region with R&D expenditures above 2.5% of its GDP, likely leading to its status as a moderate rather than emerging innovator. It is a nationally leading region when it comes to modern business services, and a major ICT centre. Other leading economic sectors include life sciences, renewable energy, chemistry, metals and machinery, and creative and leisure industries. Krakow is also an important academic hub with one of the country's most dynamic startup ecosystems, focusing on AI, business automation, and enterprise software.[31] [32] According to our analysis of Dealroom.co data, two-thirds of Polish AI startups are located in three cities: Warsaw, Krakow, and Wroclaw. With GAIA, at least one of the Polish AI factories is thus located in one of Poland's major AI startup hubs.

While the interview with an AI factory host organisation allows us to capture the supply-side perspective of the Polish AI factory ecosystem, the five interviewed AI startups represent some of the prospective users of the Polish AI factories. Therefore, they allow us to capture an important part of the demand-side perspective of the Polish AI factory ecosystem. The interviewed startups are located in 4 cities (Poznań, Warsaw, Lublin, Gliwice). They develop applied AI solutions in markets that are highly regulated and strategically important for the EU: three startups are active in the healthcare and life sciences market. They represent both an important target market for AI startups in Poland as well as Europe in general, as our recent analysis of European AI startup data gathered by Dealroom.co[33] suggests. Moreover, the healthcare market is one that the EU considers a strategic priority.[34] The two other startups target the space and defence market (one in space tech and one in defence), another strategic priority for the EU.[35] The startups interviewed thus allow us to shed light on the demand for public compute infrastructure among AI startups in applied, regulated, and strategic markets, rather than the demands of frontier AI startups who are often at the center of public discussion.

In addition to the interviews with the AI factory host as well as AI startups, we interviewed four organisations that can be considered ecosystem enablers. These

29 Marshal Office of the Wielkopolska Region, *Invest in Wielkopolska: Investor's Guidebook* (Poznań: Marshal Office of the Wielkopolska Region, 2018), 4–11.

30 European Commission, *Regional Innovation Scoreboard 2025* (Brussels: European Commission, 2025).

31 MOTIFE, *Krakow IT Market Report 2025* (MOTIFE, 2025).

32 Business in Małopolska, 'Economy', *Business in Małopolska*, n.d., accessed June 8, 2026, https://www.businessinmalopolska.pl/en/why-malopolska/economy.

33 Nicole Lemke, *European AI Competitiveness Beyond Frontier Models*, Interface, May 14, 2026, accessed May 23, 2026, https://www.interface-eu.org/publications/european-ai-competitiveness-beyond-frontier-models.

34 European Commission, 'Apply AI Strategy', *Shaping Europe's digital future*, October 2025, accessed May 26, 2026, https://digital-strategy.ec.europa.eu/en/policies/apply-ai.

35 European Commission, 'Apply AI Strategy'.

organisations provide important perspectives, because they often occupy key positions in innovation ecosystems supporting established firms and startups. In total, the organisations interviewed for this policy brief thus provide coverage across AI startups in two EU-strategic markets relevant to public-compute investment (healthcare and life sciences, defense and space) supplemented by four ecosystem/research actors, to capture the broader landscape, and one supplier-side interview (Cyfronet/GAIA AI Factory) to ground the demand-side findings against the offer being prepared. Interviews were conducted in semi-structured format, averaging 45 minutes, organised around three blocks of questions: demand for public compute; supply of public compute; and future outlook, including AI gigafactories and broader ecosystem. Analysis followed an inductive thematic approach.[36] Organisations are named and quoted with participant consent; by default, the content of the interviews is anonymised and analysed on an aggregate level.

Drawing on the interviews, the policy brief examines three challenges that condition whether public compute investments can potentially translate into commercial impact: the operational and institutional barriers that determine whether private firms can use AI factory infrastructure even when subsidised access is on offer; the actual content of private-sector demand for public compute beyond the LLM-centric framing of the policy debate; and the portability of workloads between public infrastructure, commercial cloud, and in-house systems as firms scale. The brief sets out what Polish AI startups want from public compute, what AI Factories are designed to offer, and what would need to change for the gap between the two to close.

While both AI factory projects were launched in 2025,[37] [38] neither of the facilities are yet operational. This means that our research predominantly covers expected rather than concrete demand as well as perceived rather than real requirements for the infrastructure's functionality. At the same time, this allows us to look at AI factories at the point at which design choices about access, services and pricing are still being made, in a window in which policy adjustments are most feasible, before access models, governance routines and operating tempo become entrenched.

Poland is also engaged in negotiations with the European Commission over hosting an AI gigafactory as part of a multi-state consortium, although, as mentioned earlier, the gigafactory initiative as a whole remains at a pre-call stage.[39] Compared to AI factories, which are supposed to support AI innovation and adoption across the continent, AI

36 Michelle E. Kiger and Lara Varpio, 'Thematic analysis of qualitative data: AMEE Guide No. 131', *Medical Teacher* 42, no. 8 (2020): 846–854, https://doi.org/10.1080/0142159X.2020.1755030.

37 Poznańskie Centrum Superkomputerowo-Sieciowe, 'The Launch of the First Polish AI Factory – PIAST-AI', Poznan Supercomputing and Networking Center PSNC, June 23, 2025, accessed May 15, 2026, https://www.psnc.pl/the-launch-of-the-first-polish-ai-factory-piast-ai/.

38 Academic Computer Centre Cyfronet AGH, 'Gaia AI Factory inauguration – one of the most important Europe's AI projects emerges', accessed May 15, 2026, https://www.cyfronet.pl/en/news/2026/inauguracja-gaia-ai-factory.

39 Bochyńska, 'KE zmienia terminy, Polska czeka'.

gigafactories were originally framed as explicitly targeting frontier AI development, with considerably more compute available at each location. While this is not a necessity – with recent interface research[40] also exploring multi-client models for AI gigafactories – it has shaped public perception and the debate surrounding the gigafactory initiative. Poland's position in the prospective gigafactory consortium also allows us to get a clearer picture of the needs and actors of the national AI ecosystem regarding AI gigafactories as well as the relationship between AI factories and gigafactories.

As a single country case study, the findings presented in this policy brief are naturally limited to the scope of the country selected, Poland. At the same time, there are several aspects of our research we expect to have broader applicability across other AI factory ecosystems: the operational requirements expressed by the startups we interviewed (cost, speed, IP protection, cluster isolation, portability between public compute and commercial cloud, and data residency) are not tied to the Polish context, and we therefore expect them to broadly hold across AI factories serving applied AI startups in EU-strategic sectors elsewhere in Europe.

We expect demand-side factors to fluctuate depending on the types of products AI startups develop, the development stage of the startup, the market and regulatory environment they are operating in, as well as the broader offers of the innovation ecosystem they operate in. AI factories' capability building offers are likely of greater importance in ecosystems that do not already offer highly competitive conditions for AI startups. Moreover, SMEs may have different or additional requirements for AI factories that are not mentioned in this policy brief.

On the supply side, we expect AI factories of similar types to encounter similar challenges: For example, PIAST and GAIA are both research-led consortia and the challenges regarding bridging academic and commercial demands emerging from this are likely applicable to AI factories of the same type. Beyond these characteristics, we equally expect the ecosystem AI factories are located in matter for the challenges they encounter. In the case of Poland, this includes the structure and timing of the Polish state-aid mechanism for the national co-funding share, the limited size of the Polish-language training corpus, the particular configuration of the PIAST and GAIA consortia, and the geographic distribution of Poland's AI startup hubs. Policymakers should keep these factors in mind when interpreting the findings presented in this policy brief.

---

40 Maria Nowicka, *From Chips to Grids: Why Energy Constraints Influence AI Compute Expansion in Europe*, Interface, May 5, 2026, accessed May 24, 2026, https://www.interface-eu.org/publications/ai-compute-energy-bottlenecks.

Findings

The ten interviews conducted with Polish AI startups, ecosystem actors and one factory operator surface a coherent picture: demand for public AI compute among applied European AI startups is real, sectorally concentrated in EU-strategic domains (healthcare and life sciences, space and defence), but only partially aligned with the AI factory model as currently designed. We present our findings as follows: First, we detail what applied European AI startups want from public compute. Second, we outline what AI factories currently offer. Third, we point out which challenges remain and how they can be overcome.

What AI startups want from public compute

Cost and speed: the entry conditions

Polish startups across regulated and strategic sectors for AI development see real value in subsidised public compute. All interviewed AI startups across markets, and 7 out of 10 of all respondents, name the cost of compute as an important constraint for their commercial activity. At the same time, 5 out of 10 respondents pointed towards the potential bureaucracy around accessing public compute in a timely manner as an important barrier for the usability of such infrastructure:

"Here the process has to be light enough that I submit now and within two, three, four weeks at most I have a decision: yes, I got the grant, and shortly afterwards I can get onto the cluster." - Startup 1

Operational standards

Respondents also name a consistent set of operational requirements that AI factory infrastructure must meet to make it interesting for AI startups. The first requirement is IP ownership: Six out of 10 respondents raise IP-security concerns spanning both technical and legal dimensions, e.g. because the training data is the core value of the company that needs to be protected. A second requirement is data residency: storing and processing data in Europe is mentioned by all 10 respondents, usually as a suitable use case for public compute for certain regulated sectors that require it, e.g. defence or health. The need for infrastructure located solely in the EU appears to be sector and product-dependent, not universal. Managed environments are another important requirement for respondents: 5 out of 10 prefer managed HPC environments over raw machine access. This is often mentioned as an explicit advantage public compute infrastructure could provide, with explicit guidance being an important part of the potential offer. Two additional important requirements are cluster isolation and external auditability: 3 out of 10 respondents articulate explicit requirements for cluster isolation that would withstand third-party security audit, also fearing the unwarranted access to their data.

“What could be problematic is the absence of an SLA [service level agreement: ‘contract between a service provider and a customer that defines the service to be provided and the level of performance to be expected’[41] *, or really the lack of trust in the resilience and what is actually going on there. If there were a real possibility of proper external auditing, verification of those centres and access by, say, industry audit firms like Securitum, rather than the Ministry of Digitalisation, I think that would substantially raise the potential of these places.” - Startup 2*

Lastly, respondents mention portability between AI factories, AI factories and gigafactories, and AI factories and commercial providers as crucial requirement: 4 out of 10 name portability and cloud-switching as a hard operational requirement: factory APIs that mirror commercial cloud APIs, and a container-based delivery stack built on industry standards (Docker, Kubernetes, YAML) so that subsidised public access does not turn into a new form of vendor lock-in once a product matures.

“What's interesting about the commercial clouds is that the APIs they expose are standardised. If [factory] APIs were standardised according to what the commercial clouds provide, then switching between commercial and public ones would be much easier, because it would happen on the same principles.” - Startup 3

Local human support as a positive value

Five out of 10 respondents identify locally available human support as a positive value of public compute. The argument is operational rather than sentimental, as hyperscalers’ users-per-staff ratio is a bottleneck in itself. A local operator, such as an AI factory, can offer a better one. The gain is access to consultation hours, phone access in place of ticket queues, and engineers who can talk to clients in Polish about sector-specific problems. Ecosystem and supplier interviewees alike treat this as one of the few categorical advantages a public-mission operator can sustain against commercial cloud.

“The more users these various clouds have, the harder it is to get good-quality support and a solution. […] My experience was that once I received a compute grant from a newer cloud that was just entering the market and wasn't one of the largest providers,

41 IBM, ‘What Is an SLA (service level agreement)?’, *IBM Think*, May 30, 2024, accessed June 3, 2026, https://www.ibm.com/think/topics/service-level-agreement.

and the technical support was significantly faster, by a good order of magnitude." - Research body 1

Data, not just compute

Six out of 10 respondents identify data as a fundamental bottleneck. Three issues recur: the siloed condition of Polish public-sector data, the limited size of the Polish-language corpus available for training, and firms' own bandwidth to prepare data into AI-ready form. The implication is that an AI factory whose data-hub function (explicitly part of the EuroHPC AI factory definition) is left underdeveloped relative to its compute capacity risks meeting a demand that cannot be exercised. Institutions and firms are willing in principle but not yet operationally ready; without an upstream data layer, subsidised compute supplies a tool with no input.

"It often comes down to the fact that you need data. If we don't have the data, then even with whatever hardware imaginable, we wouldn't be able to do anything with it. So I think that's probably the biggest bottleneck; preparing the data for what you actually want to do." - Startup 3

Training, inference, and sectoral specificity

The startups interviewed are concentrated in EU-strategic sectors (healthcare and life sciences, and space and defence). Among the five interviewed, four frame training as their primary use case for public compute, and one mentions inference; notably, none describes its core work as LLM-centric, the training paradigm around which much of the public-compute debate is built. These preferences appear to be tied more to product stage and type than to sector, as there was a split among the respondents from the healthcare and life sciences market. These preferences will likely shift as startups and their products mature, underlining the need to account for changing compute needs throughout startup and product life cycles. Training preference frequently correlates with R&D-heavy work where production inference either runs on the customer's edge device, migrates to the commercial cloud anyway, or has not yet been planned in detail at the firm's current stage.

This last case is itself an opportunity for AI startups: where startups have not yet decided how inference will be served, AI factories that engage earlier in firms' development planning, positioning themselves as a partner across the product lifecycle rather than only at the training stage, which could strengthen their long-term use case in Poland. This also underlines the importance of portability across AI factories, gigafactories, and commercial providers, as well as off-ramps from AI factories as pointed out by earlier interface research.[42] How prospective AI gigafactories will fit into this life cycle is another open question: the Commission's framing of the gigafactory initiative, originally pitched around training, has itself begun to shift toward explicitly including inference.[43]

Whether public compute infrastructure should offer capacities for inference, is something the respondents' opinions diverge about:

"It seems to me that Polish factories don't necessarily need to be a good place for inference. That can be left to the commercial clouds, because it's something you have to have globally anyway. We're not going to win on globalisation; you need those networks spread across the world. […] The big gaps are at the training and algorithm-development stage." - Startup 2

"If we're talking about training, there isn't such a big problem; costs are high, but over a short time horizon, a month or two. But once the model exists […], suddenly you have to run it on a machine 24 hours a day, multiplied not by two months but by a year or two, and those are really the highest costs. Not the training itself; inference is the most expensive part of all this." - Startup 3

These two accounts matter for the design of public compute infrastructure. The first citation suggests that public compute is most valuable at the training and algorithm-development stage, with inference left to commercial clouds. The second citation identifies inference as the dominant long-run cost, but the implied need is financial support for sustained inference, not necessarily the provision of public compute as the venue where that inference runs. Public infrastructure can address the first need directly; the second may call for complementary funding instruments to foster commercial AI deployment that goes beyond what the provision of public compute infrastructure alone can currently offer.

Pragmatic operational criteria for sovereign infrastructure

Half of the respondents reject the framing of public AI compute (a framing more closely associated with the gigafactory programme than with AI factories themselves) as primarily a platform for sovereign AI model training. What they want from sovereignty is a set of concrete infrastructure properties: contractual IP ownership, EEA data residency, an operator that does not commercialise the data passing through, and the guarantees stemming from independent security audits. For the applied AI startups interviewed, sovereignty is therefore a question of where the data sits, who owns the IP, and on what terms, which is a set of operational guarantees that is largely provider-agnostic. Effectively, these firms are not asking for a European provider as such; if a commercial hyperscaler could contractually deliver the same IP terms, EEA residency,

42 Lemke and Schneider, *The European Union's AI Factories*, 14.
43 Henning, 'EU recalibrates the case'.

non-commercialisation of data and independent auditability, most would likely be satisfied. This signifies a potential point of tension with the tech sovereignty package, which frames sovereignty around the provider and the model layer rather than around the operational properties applied startups require.

“We initially tried using data center services, [...] we ultimately ended up buying our own machines and hosting on-premise for two reasons: cost, which is the primary factor, and of course, IP security. [...] For me, data sovereignty is closely tied to intellectual property. For instance, for a company like ours, up to 95% of our company's value and IP is derived from the data we have collected and annotated.” - Startup 2

The model layer raises a separate question, which model sits at the foundation and whether it is European, and it is one these firms likely do not really face. The applied startups we interviewed build narrow, domain-specific systems, and none build the large, general-purpose models around which much of the sovereignty debate is organised, nor does any describe its core work as 'LLM-centric'. For firms building frontier AI or foundation models, the picture may differ, and the model layer may be where sovereignty concerns concentrate, but that is a different population from the applied firms this brief examines. We therefore do not frame our findings as a mismatch about model sovereignty or LLM-centricity. The disconnect we identify lies in the infrastructure requirements set out above, which are operational and provider-agnostic, not in which model occupies the foundation layer.

What AI factories offer

Mission and target user

Cyfronet (GAIA AI host) positions AI factories as an incubator rather than a hyperscaler competitor. The "credit-card-and-go" clientele that needs the fastest and simplest possible access is explicitly conceded to the commercial cloud. The factory's intended targets are the startups for which sovereignty is a hard constraint, regulated or IP-sensitive workloads, and startups that want to extend their runway, meaning the period during which they operate on existing cash reserves by accessing compute without selling equity to fund it. This offer fits well with the needs of the applied startups in highly regulated sectors that are its prospective users.

The explicit contrast is not with hyperscalers but with venture capital, since a factory provides early-stage compute for free without taking a stake in the company. The model the AI factory assumes has a natural exit point; once a product validates on the market, the firm migrates to commercial cloud or in-house infrastructure and the public-mission operator releases capacity for the next startup. This is the mission that the AI factory views as the core tenet of the AI factory initiative itself, not only its own model.

"Large, profitable companies are welcome too - it's not that we don't want them - but they aren't our primary target, because they already have the means to access this kind of infrastructure. What we really want is to incubate innovation in Europe, and that is the main goal of every factory. The challenge in Poland is that the venture-capital market is still underdeveloped, so startups struggle to raise capital, and where VC money is available it usually comes in exchange for equity. There are also incubation and support programmes - often EU-funded - that provide financing without taking a stake, and the European intervention works in that same spirit: we don't want to take that slice of the startup." - Marek Kasztelnik, Cyfronet, GAIA AI Coordinator

Pricing and access

In terms of pricing, GAIA aims to offer fully free access for startups and SMEs under a state-aid framework. Each AI factory is co-funded 50/50 between EuroHPC JU and the state, and the compute capacity distribution is divided accordingly. The EuroHPC half is already operable, but the national is not yet, because the corresponding domestic state-aid mechanism for awarding the Polish co-funding share has not yet been resolved. The same regulatory bottleneck can potentially emerge across other Member States hosting AI factories.

"The simple answer for all factories would be round zero - free - for startups and SMEs. But it isn't that simple, because of what's called the public-aid issue: under EU rules you can't simply give selective, subsidised access to companies, as that could be treated as aid that distorts competition. Funding for the machines is split 50/50 between the EuroHPC JU and Poland. On the EU side there is an act that exempts HPC from those rules, so for the EuroHPC half - half of every factory - there is no issue: it can already be done today, and there it is round zero. For the other half, the national share, it comes down to each country's own regulations." - Marek Kasztelnik, Cyfronet, GAIA AI Coordinator

The revised EuroHPC Regulation extends a similar access logic to gigafactories. It provides that a share of Union access time to AI factories and gigafactories shall be used 'primarily' to grant start-ups and SMEs access for research or innovation activities,[44] and that such access is to be free of charge for public-law bodies and for the private innovation activities of SMEs and scale-ups.[45] The free-for-startups principle Cyfronet describes is therefore not specific to AI factories; it is written into the legal framework for both pillars, even as the national co-funding bottleneck remains unresolved.

44 Council of the European Union, 'Council Regulation (EU) 2026/150', art. 12b(10).
45 Council of the European Union, 'Council Regulation (EU) 2026/150', art. 12b(15)(b).

Service, security, isolation

On the technical side, Cyfronet's articulation of factory security rests on certifications and architectural isolation already in place at the host entity, not on infrastructure to be built later. The host entity operates with ISO 27001[46] and TISAX[47] certifications and treats security by design as a baseline obligation that follows from its broader work with public-sector clients, not a factory-specific feature. Compute and data isolation are HPC-grade: each tenant receives negotiated compute capacity and a fully separated storage space, with dedicated management procedures. Regarding IP, the operator takes no share in IP produced on its infrastructure. The remaining open question is not whether these guarantees can be delivered, but whether they will be codified in standard access agreements so that prospective users can rely on them *ex ante*.

"We have established expertise and a proven track record on security. We hold ISO 27001 and TISAX, and the number of certifications will only grow over time - so this is security by design. As for guaranteeing that another user on the infrastructure cannot reach your environment, we have mechanisms for exactly that: isolation in both compute access and data access, validated by the certifications we hold. And as an organisation, we have absolutely no intention of taking a share of the IP produced on our infrastructure. That is the standard arrangement we offer our clients - they keep full ownership so they can develop their work freely. Joint R&D projects are a separate case, where IP is handled by agreement between the partners." - Marek Kasztelnik, Cyfronet, GAIA AI Coordinator

Training and inference; cross-factory data lab; sectoral niches

The host entity does not see the AI factory as exclusively for training or inference. The split between training and inference tracks the user's stage and the innovativeness of its solution: less ambitious firms that build on top of existing open-weight models need inference, embedding services, and vector stores; more ambitious startups in domain niches train their own models because suitable ones do not exist on the market.

"*The split between inference and training really comes down to company readiness and the innovativeness of the solution. […] We already have clients on our infrastructure where we provide an inference service and they build applications on top of an existing*

46 ISO/IEC 27001 is the international standard for information-security management systems. Certification by an accredited body signals adherence to recognised best practice. See: https://www.iso.org/standard/27001.

47 TISAX (Trusted Information Security Assessment Exchange) is an information-security assessment and exchange mechanism, based on ISO/IEC 27001/27002 with sector-specific requirements; it lets audited organisations share standardised assessment results with partners. See: https://enx.com/en-US/TISAX/.

inference service with open-weight models, we can do that, and it matters to them. But more ambitious startups do train their own solutions, particularly in the domains we focus on, so health and space technologies, where we need specialised models that we won't get off the market and have to train ourselves." *- Marek Kasztelnik, Cyfronet, GAIA AI Coordinator*

This openness towards offering both training and inference is somewhat contrasting with the startups' own assessment of the AI factory's offer as mainly a training ground for prototyping. A primary question emerging from this finding is where AI factories should focus their efforts, assuming that their compute capacity is ultimately limited, to avoid supply-side bottlenecks. Another open question is whether the AI factory model will have to shift from training to inference in the future if demand for inference exceeds demand for training.

Regarding the data bottleneck some AI startups have identified, the answer the AI factory provides is the cross-factory data lab in formation across EuroHPC AI factories. For example, Cyfronet has for years curated imagery from the Copernicus Sentinel (Earth observation satellite programme), which generates data on land, oceans and the atmosphere.[48] The point illustrates the data-lab logic: rather than asking each incoming startup to assemble a training corpus from scratch, the factory supplies a domain-relevant dataset (here, directly useful for the space and environmental use cases GAIA targets) as a starting point.

In terms of sectoral niches, the two Polish factories illustrate a broader design choice between a generalist and a more specialised AI-factory offer. GAIA is comparatively sector-focused: its announced priorities are healthcare, space technologies and large language models. PIAST advertises a broader spread, organised around clusters spanning technology & digitalisation, health & society, environment & industry and the exact sciences. Neither model is inherently superior: a specialised factory can in principle offer pre-curated data, domain expertise and tailored support that a generalist cannot match, while a generalist factory serves a wider base and hedges against demand uncertainty. The relevant question for operators is which model fits the demand profile of their ecosystem. Given finite compute, whether spreading capacity across many sectors risks under-serving the specialised, hard-to-source workloads (such as domain-specific model training) that commercial clouds are least likely to provide. For startups, the practical implication is that the two factories offer partly complementary specialisations within the same national programme, relevant when deciding where a given workload best fits.

Table 1. Declared specialisations of Poland's two AI Factories [49]

48 Dávid D. Kovács et al., 'Copernicus Data Space Ecosystem establishes public cloud processing for earth observation data', *Scientific Data* 13, no. 537 (2026), https://doi.org/10.1038/s41597-026-06765-8.
49 See: https://gaia.plgrid.pl/ & https://www.pcss.pl/piast-ai/.

	GAIA AI Factory (Kraków)	PIAST-AI Factory (Poznań)
Operating entity	Cyfronet	PSNC
Focus	Comparatively sector-focused	Comparatively generalist-focus
Focus areas	**1. Healthcare** **2. Space Technologies** **3. Large language models**	**1. Technology & digitalisation:** • cybersecurity, • virtual worlds, • quantum communication and IT. **2. Health & society:** • life sciences, • culture, • education, • humanities. **3. Environment & industry:** • smart city, • agriculture, • energy efficiency, • aviation. **4. Exact sciences:** • astronomy and physics, • materials science, • research data processing

Local support and sovereignty as property - differentiators

Local human support is treated by Cyfronet as a structural advantage stemming from a smaller users-per-staff ratio: paid consultation hours, phone access, and in-city engineer contact, against the difficulty of reaching equivalent support from big commercial providers. The second is sovereignty articulated as a property of the operator rather than of the model. Cyfronet's commitment is that data passing through the infrastructure is not commercialised, that residency stays inside the EEA by default, and that the operator's mission is public rather than commercial, which is a major difference from the business model of commercial cloud.

"When I have a company in Kraków and there's a Cyfronet in Kraków, being able to meet, or even buy consultation hours with us, or sit down and talk through problems on the infrastructure, pick up the phone and call, that is so much easier than… well, anyone who has tried to contact commercial cloud support knows exactly what that looks like. […] In our case we focus more on interaction with the client" - Marek Kasztelnik, Cyfronet, GAIA AI Coordinator

Challenges and how to make public compute work

Public aid - a European problem?

The first challenge is regulatory rather than technological. While the EuroHPC half of factory funding can already be granted at zero cost, the national half depends on a domestic state-aid mechanism that is not yet resolved in Poland. On the supplier's account, this is unlikely to be a Polish-specific issue. Confirming the scope of the problem across Member States would require comparative work beyond this single-country case study.

This unresolved national aid question also has direct governance consequences. It conditions who, in practice, decides which startups access subsidised public compute, and on what terms. Our interviews did not provide evidence on the current demand-to-capacity ratio, and the Polish factories are not yet operational at scale; the question of whether demand for free access will outstrip available compute, and how access decisions will be made when it does, therefore, remains open and warrants dedicated follow-up research.

Moreover, until that mechanism is settled, the headline promise of free access for startups and SMEs cannot be operationalised at full scale, regardless of how ready the infrastructure itself is. This is a coordination problem that no individual factory consortium can resolve. The Commission, together with national governments, could develop an EU-level template for the national half of the state-aid mechanism, covering, e.g., eligible activities and audit and verification procedures, to establish uniform rules across jurisdictions. The same logic extends to multi-state consortia of gigafactories (such as the prospective Baltic AI Gigafactory), where divergent national interests and state-aid regimes would need to be reconciled in a shared governance framework before subsidised access is granted.

Constraints in national funding could lead to differential outcomes across AI factories located in different countries, thereby potentially undermining the effectiveness of the AI factory initiative at large. This may especially be the case if the funding constraints are tied to AI factories located in moderate or emerging innovation settings, as funds may then be lacking precisely in the environments where capability building is most relevant.

Bridging academic and commercial operating cultures

The second challenge is structural. Startups expect operational standards drawn from commercial cloud norms. The supplier side confirms the underlying capability (security by design) so the gap is not in what the factories can do but in how they operate around the core service. Partnership agreements, response times, and the visibility of service standards may currently inherit the academic operating tempo from EuroHPC's research roots. A possible solution is to publish minimum service standards before each facility opens to commercial users (access tempo, isolation and audit terms, IP terms, incident response) with separate reporting tracks for commercial and academic users.

Codifying sovereignty guarantees in access agreements

Respondents expect sovereignty articulated as a set of infrastructure properties: contractual IP ownership, EEA data residency, an operator that does not commercialise data, and access to independent security audits. The host entity confirms each of these as its existing practice in principle. These commitments, however, should be both communicated to the prospective users and present in standardised access agreements that a startup user can access *ex ante*.

Portability, interoperability, cloud switching

The fourth challenge concerns the natural transition from a factory to a commercial cloud or in-house infrastructure once a product validates on the market, and the conditions under which that transition is actually feasible. At the European level, the EuroHPC Federation Platform launched in April 2026 as a unified access point across EuroHPC HPC systems, with AI factory integration planned for subsequent releases.[50] [51] The challenge is non-trivial even at this internal level, since European HPC sites currently run on incompatible hardware-software stacks (for example, AMD ROCm at Finland's LUMI vs Nvidia CUDA at sites such as Spain's MareNostrum 5).[52] Relatedly, recent research has argued that EuroHPC AI factories are built on HPC systems optimized for raw performance but less appropriate for cloud-native usability and inference,[53] which has direct implications for the offer AI factories can realistically make.

Just as factory-to-factory compatibility is being addressed, the consortia should commit to portability between AI factories and commercial cloud providers, an argument already developed by previous interface research.[54] At a minimum, this should mean publishing portability and interoperability commitments alongside technical-level harmonisation

50 European High Performance Computing Joint Undertaking, 'EuroHPC Federation Platform', accessed May 23, 2026, https://www.eurohpc-ju.europa.eu/supercomputers/eurohpc-federation-platform_en.

51 European High Performance Computing Joint Undertaking, 'First release of the EuroHPC Federation Platform to streamline access to Europe's supercomputing resources', European High-Performance Computing Joint Undertaking, April 15, 2026, accessed May 23, 2026, https://www.eurohpc-ju.europa.eu/first-release-eurohpc-federation-platform-streamline-access-europes-supercomputing-resources-2026-04-15_en.

52 Kyosovska and Renda, *EU plans for AI (giga)factories*, 33–34.

53 Pedro Garcia Lopez et al., 'AI Factories: It's time to rethink the Cloud-HPC divide', arXiv, September 16, 2025, https://arxiv.org/abs/2509.12849.

54 Lemke and Schneider, *The European Union's AI Factories*, 14.

across the EuroHPC network, so that workloads can move between factories, and between factories and commercial cloud, once products mature. Factory-to-factory compatibility is not only a technical convenience but a single-market question: a genuinely interoperable network lets compute, models and expertise circulate across member states, turning nineteen national facilities into shared European capacity rather than isolated nodes. Second, building clear pathways to scale, including off-ramps from public infrastructure to commercial providers,[55] allows successful European startups to grow without leaving the continent for non-EU hyperscalers, keeping data, IP and economic value within the EU as firms expand. Lastly, interoperability should be a key requirement for the infrastructure yet to be built at certain AI factory locations as well as future updates to the infrastructure hosted at existing locations.

The upstream data layer

The data bottleneck identified by respondents is partly a national governance problem: the siloed condition of public-sector data, the limited size of the Polish-language corpus, and firms' own bandwidth to prepare data into AI-ready form are challenges which no single factory consortium can resolve. The cross-factory data lab in formation across EuroHPC AI factories is a promising contribution, and one already reflected in the European Commission's Data Union Strategy, which identifies AI Factory Data Labs as the link between EU data assets and AI infrastructure.[56]

On the supplier's account, the cross-factory data lab is at the working-group stage across factories, a natural early phase that should progress to a funded programme line with budget, deliverables, and a public roadmap on the same footing as the compute investment. This should run alongside coordinated efforts with national digital ministries on a parallel public-data desiloing track, so that compute capacity and AI-ready data arrive around the same time. Without this pairing, AI factories will continue to meet a demand that, on the data side, cannot yet be fully met.

Conclusion

The central question we addressed in this policy brief is for whom, in what form, and under what conditions can public compute support European AI innovation. The insights presented are based on ten semi-structured interviews with actors from the Polish AI innovation ecosystem: one AI factory host, four ecosystem enablers, and five Polish AI

55 Lemke and Schneider, *The European Union's AI Factories*, 14.
56 European Commission, 'Communication from the Commission to the European Parliament and the Council: Data Union Strategy – Unlocking Data for AI', COM(2025) 835 final, November 19, 2025, 7–8, accessed May 23, 2026, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52025DC0835.

startups operating in EU-strategic markets (such as healthcare and life sciences or space and defence). The interviews confirm that the demand for public compute is real, sector-specific, and that AI factories are well-placed to support local AI innovation across applied markets in Europe.

The AI startups we interviewed require a set of concrete properties from public AI infrastructure: They welcome no-cost access to HPC infrastructure, as long as this access can be provided in an unbureaucratic and time-sensitive manner, protects their data and IP, and allows for easy switching between different AI factories as well as between AI factories and commercial cloud providers. For respondents in regulated sectors, this extends to data residency, namely the ability to store and process data within the EU. They see the localised human support AI factories plan to provide as a real asset, an offering beyond that of private cloud providers. Access to data – another service AI factories are planning to provide – is equally needed.

The remaining challenges facing AI factories today – securing the national portion of the investment, bridging academic and commercial operating cultures, codifying sovereignty guarantees in access agreements, enabling portability, interoperability, cloud switching, and providing data access - are real but not insurmountable. If policymakers and AI factories are aware of these realities and tailor their services accordingly, they are well placed to translate public investment in AI compute into value added for European AI innovation.

A related question, one which factory operators and policymakers will need to confront as the infrastructure comes fully online, concerns the governance of access itself. If demand for free or subsidised public compute outstrips the capacity any single factory can offer, who decides which startups are admitted, and on what basis? Our research did not surface evidence on the current demand-to-capacity ratio: the Polish factories are not yet operational at scale, and the national share of access remains contingent on the unresolved state-aid mechanism. We therefore cannot answer this question empirically from the present case, but we flag it as a critical design choice for the AI factory programme.

When it comes to the planned AI gigafactories, our respondents largely treat them as a capacity upgrade rather than a qualitatively different category, and gigafactory access is not a stated priority across the lifecycle of their current products. The operational and technical findings of this brief (managed environments, isolation and audit, portability, codified sovereignty guarantees, local human support), therefore, likely extend to gigafactories under broadly the same conditions. The policy challenges concerning gigafactories sit at the system level and are foundational for the long-term value of the investment. One of these potential constraints is energy, which was analysed in a recent interface paper.[57]

This policy brief provides a single-country case study using limited qualitative data: ten semi-structured interviews across one AI factory host, four ecosystem enablers, and five AI startups. Nevertheless, we believe the Polish case to be instructive for understanding how public AI infrastructure investment can have a positive impact on commercial AI innovation in Europe, precisely because Poland combines significant AI factory funding with an ecosystem that does not, by conventional indicators, appear to be an obvious destination for frontier AI innovation. If public compute can stimulate commercial activity here, the model is likely transferable.

Critics of European public compute investment often generalise from the trajectory of frontier-LLM developers to the entire universe of AI firms. Our research focuses on the compute needs outside of this small and specialised population. Applied AI in regulated and strategic sectors creates real demand for European compute – under conditions that may differ from what frontier model training requires. The startups we interviewed need flexible, unbureaucratic access to HPC infrastructure that protects their data and IP, enables portability between providers, and is backed by localised human support. These are achievable design choices, not structural obstacles.

The central implication is not whether to invest in public compute, but how to design it. AI factories that tailor their services to the operational realities of applied AI startups – rather than exclusively optimising for a frontier training paradigm that few European firms are pursuing – are well placed to translate public investment into durable value for European AI innovation.

Acknowledgements

We are grateful to the many individuals and organisations who made this publication possible. We thank, first and foremost, the respondents and organisations who generously volunteered their time and expertise to support our research. We are also indebted to our colleagues Catherine Schneider and Maria Nowicka for their valuable feedback on earlier drafts, and to Luisa Seeling for her careful editing and support throughout the publication process. Thanks are also due to Alina Siebert and Michael Brönner for making the publication available online, and to Iana Pervazova for helping to disseminate our findings. Finally, we thank Francesco D'Alessandro and our other peer reviewers, whose expert input has strengthened this work considerably.

57 Nowicka, *From Chips to Grids: Why Energy Constraints Influence AI Compute Expansion in Europe*.

Bibliography

Academic Computer Centre Cyfronet AGH. ‘Gaia AI Factory’. Accessed May 14, 2026. https://gaia.plgrid.pl/.

Academic Computer Centre Cyfronet AGH. ‘Gaia AI Factory inauguration – one of the most important Europe’s AI projects emerges’. Accessed May 15, 2026. https://www.cyfronet.pl/en/news/2026/inauguracja-gaia-ai-factory.

Badowski, Łukasz. ‘Polskie startupy zarabiają miliony na AI. Technologia made in Poland podbija świat’. *INNPoland*, March 12, 2026. Accessed May 20, 2026. https://innpoland.pl/222118,polskie-startupy-zarabiaja-miliony-na-ai-technologia-made-in-poland-podbija-swiat.

Batko, Tomasz. ‘W trzyetapowym konkursie Komisji Europejskiej (KE) zdecydowano o utworzeniu 19 Fabryk AI w 16 krajach Unii Europejskiej’. LinkedIn, 2026. Accessed May 14, 2026. https://www.linkedin.com/feed/update/urn:li:activity:7416928301056819200/.

Bertuzzi, Luca, and Lewis Crofts. ‘EU’s AI gigafactories project plagued by tech and budgetary bottlenecks’. *MLex*, April 13, 2026. Accessed May 23, 2026. https://www.mlex.com/mlex/articles/2463944/eu-s-ai-gigafactories-project-plagued-by-tech-and-budgetary-bottlenecks.

Bochyńska, Nikola. ‘KE zmienia terminy, Polska czeka. Niepewność wokół gigafabryki AI’. *WNP.pl*, May 14, 2026. Accessed May 15, 2026. https://www.wnp.pl/tech/ke-zmienia-terminy-polska-czeka-niepewnosc-wokol-gigafabryki-ai,1061717.html.

Business in Małopolska. ‘Economy’. *Business in Małopolska*, n.d. Accessed June 8, 2026. https://www.businessinmalopolska.pl/en/why-malopolska/economy.

Council of the European Union. ‘Council Regulation (EU) 2026/150 of 16 January 2026 amending Regulation (EU) 2021/1173 on establishing the European High Performance Computing Joint Undertaking’. *Official Journal of the European Union* L, 2026/150 (January 19, 2026). http://data.europa.eu/eli/reg/2026/150/oj.

European Commission. ‘Apply AI Strategy’. *Shaping Europe’s digital future*, October 2025. Accessed May 26, 2026. https://digital-strategy.ec.europa.eu/en/policies/apply-ai.

European Commission. ‘Communication from the Commission to the European Parliament and the Council: Data Union Strategy – Unlocking Data for AI’. COM(2025) 835 final, November 19, 2025. Accessed May 23, 2026. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52025DC0835.

European Commission. ‘EU launches InvestAI initiative to mobilise €200 billion of investment in artificial intelligence’. *Shaping Europe’s digital future*, February 11, 2025. Accessed May 14, 2026. https://digital-strategy.ec.europa.eu/en/news/eu-launches-investai-initiative-mobilise-eu200-billion-investment-artificial-intelligence.

European Commission. ‘European Innovation Scoreboard 2025: Country profile Poland’. Brussels: European Commission, July 2025.

European Commission. *Regional Innovation Scoreboard 2025*. Brussels: European Commission, 2025.

European High Performance Computing Joint Undertaking. ‘AI Factories’. Accessed May 12, 2026. https://www.eurohpc-ju.europa.eu/ai-factories_en.

European High Performance Computing Joint Undertaking. ‘EuroHPC Federation Platform’. Accessed May 23, 2026. https://www.eurohpc-ju.europa.eu/supercomputers/eurohpc-federation-platform_en.

European High Performance Computing Joint Undertaking. ‘First release of the EuroHPC Federation Platform to streamline access to Europe’s supercomputing resources’. April 15, 2026. Accessed May 23, 2026. https://www.eurohpc-ju.europa.eu/first-release-eurohpc-federation-platform-streamline-access-europes-supercomputing-resources-2

Garcia Lopez, Pedro, Daniel Barcelona Pons, Marcin Copik, Torsten Hoefler, Eduardo Quiñones, Maciej Malawski, Peter Pietztuch, Thomas Ohlson Timoudas, Aleksander Slominski, and Alberto Marti. ‘AI Factories: It’s time to rethink the Cloud-HPC divide’. arXiv, September 16, 2025. https://arxiv.org/abs/2509.12849.

Haeck, Pieter. ‘EU accused of wasting €20B on AI computing dreams’. *Politico*, May 4, 2026. Accessed May 14, 2026. https://www.politico.eu/article/eu-accused-wasting-20-billion-euro-ai-computing-dreams/.

Henning, Maximilian. ‘EU recalibrates the case for AI gigafactories’. *Euractiv*, April 7, 2026. Accessed May 23, 2026. https://www.euractiv.com/news/eu-recalibrates-the-case-for-ai-gigafactories/.

Hess, Julia Christina, and Felix Sieker. *Built for Purpose? Demand-Led Scenarios for Europe’s AI Gigafactories*. Policy Brief. Interface and Bertelsmann Stiftung, October 22, 2025. Accessed May 23, 2026. https://www.interface-eu.org/publications/ai-gigafactories.

IBM. ‘What Is an SLA (service level agreement)?’. *IBM Think*, May 30, 2024. Accessed June 3, 2026. https://www.ibm.com/think/topics/service-level-agreement.

Kiger, Michelle E., and Lara Varpio. ‘Thematic analysis of qualitative data: AMEE Guide No. 131’. *Medical Teacher* 42, no. 8 (2020): 846–854. https://doi.org/10.1080/

0142159X.2020.1755030.

Kögl, Larissa. 'Aus fünf mach sieben: EU denkt KI-Gigafactories neu'. *PoliticoPro*, May 28, 2026. Accessed June 8, 2026. https://pro.politico.eu/news/aus-fuenf-mach-sieben-eu-denkt-ki-gigafactories-neu.

Kovács, Dávid D., Jan Musial, Jędrzej Bojanowski, Dennis Clarijs, Jurry de la Mar, and András Zlinszky. 'Copernicus Data Space Ecosystem establishes public cloud processing for earth observation data'. *Scientific Data* 13, no. 537 (2026). https://doi.org/10.1038/s41597-026-06765-8.

Kyosovska, Nicoleta, and Andrea Renda. *EU plans for AI (giga)factories: sanctuaries of innovation, or cathedrals in the desert?*. CEPS In-Depth Analysis. Centre for European Policy Studies, November 2025. Accessed May 12, 2026. https://www.ceps.eu/ceps-publications/eu-plans-for-ai-gigafactories-sanctuaries-of-innovation-or-cathedrals-in-the-desert/.

Lemke, Nicole. *European AI Competitiveness Beyond Frontier Models*. Interface, May 14, 2026. Accessed May 23, 2026. https://www.interface-eu.org/publications/european-ai-competitiveness-beyond-frontier-models.

Lemke, Nicole, and Catherine Schneider. *The European Union's AI Factories: Lessons for Public Investment in AI Infrastructure in Europe*. Policy Brief. Interface, October 30, 2025. Accessed May 15, 2026. https://www.interface-eu.org/publications/ai-factories.

Marshal Office of the Wielkopolska Region. *Invest in Wielkopolska: Investor's Guidebook*. Poznań: Marshal Office of the Wielkopolska Region, 2018.

MOTIFE. *Krakow IT Market Report 2025*. MOTIFE, 2025.

Nowicka, Maria. *From Chips to Grids: Why Energy Constraints Influence AI Compute Expansion in Europe*. Interface, May 5, 2026. Accessed May 24, 2026. https://www.interface-eu.org/publications/ai-compute-energy-bottlenecks.

PFR Ventures, and Inovo.vc. *Polish VC Market Outlook Q1 2026*. PFR Ventures; Inovo.vc, 2026.

Pinheiro, Flávio L., Pierre-Alexandre Balland, Ron Boschma, and Dominik Hartmann. 'The dark side of the geography of innovation: relatedness, complexity and regional inequality in Europe'. *Regional Studies* (2022). https://www.tandfonline.com/doi/full/10.1080/00343404.2022.2106362.

Poznańskie Centrum Superkomputerowo-Sieciowe. 'PIAST-AI'. Accessed May 14, 2026. https://www.pcss.pl/piast-ai/.

Poznańskie Centrum Superkomputerowo-Sieciowe. ‘The Launch of the First Polish AI Factory – PIAST-AI’. Poznan Supercomputing and Networking Center PSNC, June 23, 2025. Accessed May 15, 2026. https://www.psnc.pl/the-launch-of-the-first-polish-ai-factory-piast-ai/.

Rutkowska, Elżbieta. ‘Polska może stać się hubem AI dla całego kontynentu. Wszystko zależy od tej decyzji’. *WNP.pl*, May 7, 2026. Accessed May 23, 2026. https://www.wnp.pl/tech/polska-moze-stac-sie-hubem-ai-dla-calego-kontynentu-wszystko-zalezy-od-tej-decyzji,1057662.

Authors

Jan Króliński
Fellow AI Systems, Markets & Governance
jan.krolinski@gmail.com

Dr. Nicole Lemke
Senior Policy Researcher AI Systems, Markets & Governance
nlemke@interface-eu.org

Imprint

interface – Tech analysis and policy ideas for Europe
(formerly Stiftung Neue Verantwortung)

W www.interface-eu.org
E info@interface-eu.org
T +49 (0) 30 81 45 03 78 80
F +49 (0) 30 81 45 03 78 97

interface – Tech analysis and policy ideas for Europe e.V.
c/o Publix
Hermannstraße 90
D-12051 Berlin

Design by Make Studio
www.make.studio
Code by Convoy
www.convoyinteractive.com