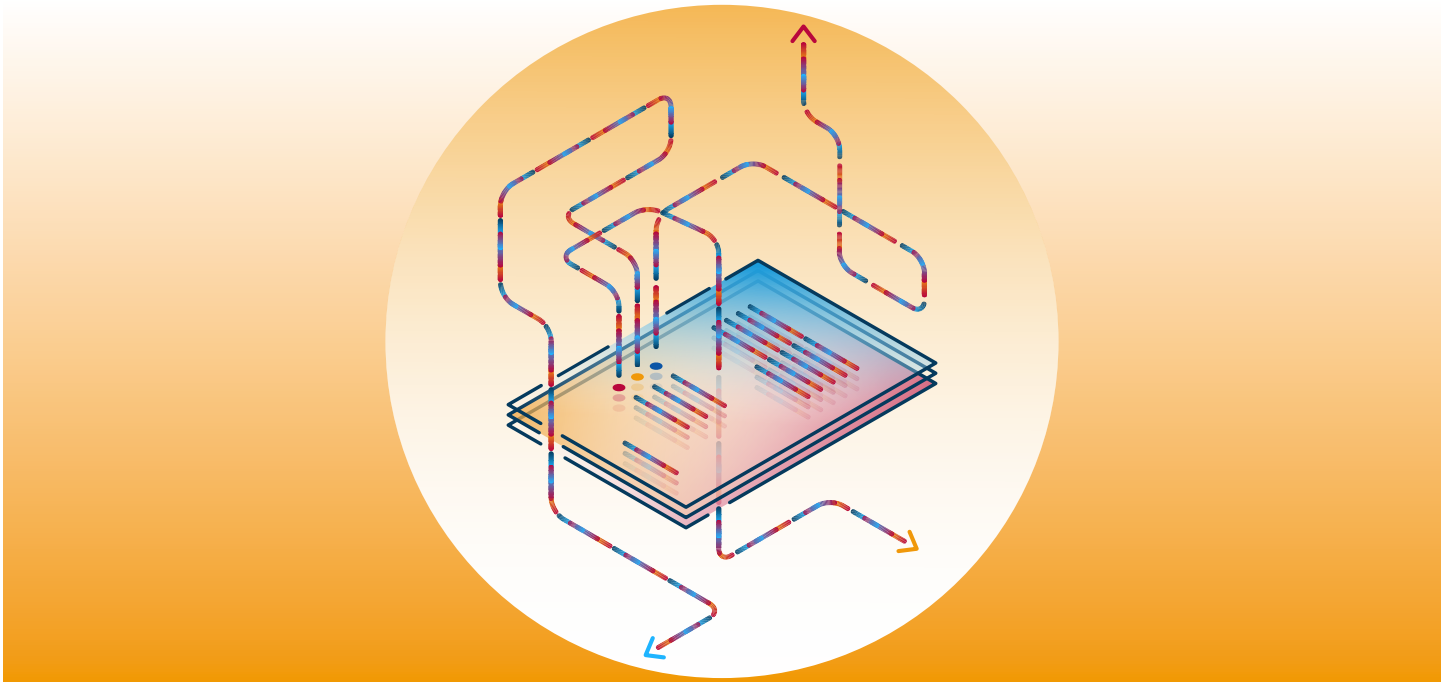

January 2026

European Public AI – Policy Brief

Dr. Alek Tarkowski, Dr. Felix Sieker

The current AI landscape is shaped by a massive gravitational pull toward capital-intensive, proprietary scaling. This dynamic has driven an unprecedented concentration of power, with control over the entire technological life cycle – from raw compute and datasets to final applications – consolidated in the hands of a small number of non-European private actors. AI development is also increasingly framed as a race, driven by competition between two superpowers, the United States and China, and a narrow group of companies based in those two countries.

The European Union has responded with the AI Continent Action Plan and a range of related strategies and programs, including Invest AI, Apply AI and the Data Union Strategy. If Europe enters this race without articulating a sovereign logic of AI development, however, it risks spending billions while reinforcing dependence on the very dependence on proprietary ecosystems it seeks to challenge. A rapid push to deploy AI solutions without sovereign foundations risks building European services on infrastructure that Europe does not control. Here, "sovereignty" should be understood as resilience and the ability to sustain critical capabilities under stress – not as technological autarky or a drive to build everything domestically.

This policy brief proposes an alternative path: a public AI strategy that treats European AI as public digital infrastructure. Such a mission-driven approach would ensure that the entire AI life cycle – from compute investments in AI factories to the deployment of ApplyAI sectoral flagships – supports market competition and innovation while also advancing the public interest and democratic resilience.

This represents a shift from a supply-driven AI race to a demand-driven approach centered on the needs of European citizens, communities and organizations. It requires a fundamental reassessment: AI is not a self-justifying solution, but a tool the value of which must be measured against its ability to meet documented public needs.

OPEN _FUTURE

BertelsmannStiftung

We argue that existing initiatives, while extensive, are also fragmented and aimed largely at facilitating private market entry. Regulation alone will not resolve this challenge. The European Union must actively support the development of a family of open-source foundation models, deployed under democratic control, to serve as a shared foundation for AI applications in the public interest. These models should form the core of an integrated public AI stack, complemented by measures that secure the underlying compute, data and software as shared, resilient commons for all Europeans.

AI as normal technology

Public debate around artificial intelligence often swings between uncritical techno-optimism and alarmist warnings about existential risks. Some developers claim artificial general intelligence is less than a decade away. These polarized narratives, however, tend to obscure how AI systems already operate in practice.

This paper adopts a more grounded perspective, treating AI as a "normal technology." While acknowledging its potential for far-reaching and gradual societal effects, this approach centers on the tangible systemic risks and infrastructural dependencies AI creates today. Public AI strategy should rest on a clear-eyed assessment of the technology's current limits and its role as a sovereign, public-interest tool for Europe. Viewing AI as a normal technology means moving away from hype and toward public value.

The term "AI" itself is imprecise, covering a broad range of computational methods. In this policy brief, the focus is primarily on generative AI systems, including large language and multimodal models, that underpin the current wave of innovation. The proposal also encompasses specialized AI models developed for specific industrial, scientific or public-sector uses. Addressing both strands, a public AI strategy aims to build a European AI ecosystem that functions as public digital infrastructure.

1 | Concentration of power in AI as a major risk

From a European perspective, the concentration of power in AI is particularly stark. Of more than 30 of the world's most capable AI models, only two have been developed by European companies.[1, 2] Almost all leading AI firms are based in the United States and China – countries that also lead in AI-related investment, talent and research and development capacity.

Mario Draghi's report on European competitiveness identifies structural weaknesses in Europe's technology sector as a central driver of the productivity gap between the EU and the United States. Today, about 80% of the digital products, services and infrastructure used in Europe are provided by non-EU companies.[3] A recent study illustrates the scale of this dependency: annual European spending on cloud and software services totals €264 billion, with 83% flowing to U.S. providers.[4] Three U.S. "hyperscalers" account for more than 65% of the European, and global, cloud market. These companies increasingly control multiple critical layers of the AI stack, including semiconductors, cloud infrastructure, models and developer platforms,[5] and they are building power through in-

1 Epoch AI, "Over 30 AI Models Have Been Trained at the Scale of GPT-4," Epoch AI Data Insights, January 30, 2025, https://epoch.ai/data-insights/models-over-1e25-flop

2 One of the most widely cited benchmarks for assessing the quality of frontier models comes from Epoch AI, which estimates the training compute used for leading models. Epoch AI defines "large models" as those trained using more than 10^25 floating-point operations (FLOP). FLOPs are the standard metric for computing power, quantifying the number of calculations a processor can perform, typically measured per second. The first model trained at this scale was GPT-4, released by OpenAI in March 2023. By June 2025, 33 AI models had crossed this threshold. Of these, only two were developed in Europe, both by Mistral.

3 European Parliament, Report on European Technological Sovereignty and Digital Infrastructure, A10-0107/2025 (Brussels: European Parliament, June 11, 2025), https://www.europarl.europa.eu/doceo/document/A-10-2025-0107_EN.html

4 Cigref, La dépendance technologique aux softwares & cloud services américains : une estimation des conséquences économiques en Europe (Paris: Cigref, April 25, 2025), https://www.cigref.fr/la-dependance-technologique-aux-softwares-cloud-services-americains-une-estimation-des-consequences-economiques-en-europe

5 Max von Thun and Daniel A. Hanley, Stopping Big Tech from Becoming Big AI: A Roadmap for Using Competition Policy to Keep Artificial Intelligence Open for All (Washington, DC: Open Markets Institute, 2024), https://www.openmarketsinstitute.org/publications/report-stopping-big-tech-big-ai-roadmap

vestments.[6] In 2023, two-thirds of the $27 billion raised by emerging AI companies came from Big Tech companies.[7]

This gap is widening, driven by unprecedented levels of private investment in cloud and AI technologies. In 2024, corporate investment in AI reached $252 billion, a 26% increase from the previous year. Investment is heavily skewed toward the United States, which accounted for $109 billion in private AI investment, compared with $9 billion in China and a combined $19 billion across EU member states.[8] In early 2025, OpenAI, Oracle and SoftBank announced a $500 billion investment in cloud infrastructure, followed shortly by similar commitments from other major cloud and AI providers.

For Europe, these developments pose both an economic risk and a challenge to the EU's strategic autonomy, democratic principles and social model. The issue goes beyond market dominance or research leadership by non-EU actors. When a small number of private providers control core AI infrastructure and key foundation models, and credible public or European alternatives are lacking, governments, regulators, public institutions, civil society and European businesses become structurally dependent on their technical choices, pricing and business models. These offerings are designed to maximize shareholder value, regardless of deployment context or public interest.

European companies and public institutions remain dependent on foreign corporations at every stage of the generative AI pipeline, from access to advanced chips and large-scale cloud infrastructure to the training, fine-tuning and deployment of foundation models. More broadly, the EU's reliance on dominant digital platforms for critical infrastructure across all layers of the internet stack exposes it to the risk of strategic dependency. Developing sovereign AI infrastructure will be essential to reducing these dependencies.

Sovereignty in this context does not mean technological autarky, but resilience. Given the depth of global interdependence in AI supply chains, full sovereignty is neither realistic nor desirable. Instead, resilience implies reducing critical vulnerabilities while deepening cooperation with like-minded partners. In practice, this could mean ensuring that compute resources for critical sectors such as defense or health care are physically located within the EU and operated by European companies, even if underlying components continue to be sourced from outside the bloc.

Reliance on non-European services is therefore more than an economic concern, and the concentration of power in cloud and AI markets is not simply a matter of market share or technological capability. Generative AI and cloud infrastructure are pervasive. They underpin everyday services such as search, navigation and communication; they are increasingly embedded in public administration, education and health care; they shape security-sensitive systems, including defense, energy grids, transport and telecommunications; and they influence information flows, media and political debate.

When control over such essential infrastructure rests with a narrow group of foreign companies, this creates risks for Europe's security and resilience, its ability to innovate on its own terms, its environmental objectives, given the resource intensity of AI and data centers, and the protection of its social model and fundamental rights. These concerns are reflected, among other sources, in the recent Council Conclusions on European Competitiveness in the Digital Decade.[9]

2 | Defining public AI

The rapid concentration of power in the AI market calls for a fundamental shift toward public AI. This policy framework treats artificial intelligence as essential public digital infrastructure and encompasses

6 Nathan Kim and David Gray Widder, "'Big Cloud' Is Building Power via Pervasive Investments," TechPolicy.Press, August 12, 2025, https://www.techpolicy.press/big-cloud-is-building-power-via-pervasive-investments/

7 George Hammond, "Big Tech outspends venture capital firms in AI investment frenzy", December 29, 2023, https://www.ft.com/content/c6b47d24-b435-4f41-b197-2d826cce9532

8 Stanford University Human-Centered AI, AI Index Report 2025 (Stanford, CA: Stanford University, 2025), https://hai.stanford.edu/assets/files/hai_ai_index_report_2025.pdf

9 Council of the European Union, ST-16430-2025-INIT (Brussels: Council of the European Union, 2025), https://data.consilium.europa.eu/doc/document/ST-16430-2025-INIT/en/pdf

policy interventions that support the creation of alternatives to dominant commercial AI systems.

Unlike these proprietary models, public AI is defined by three core pillars designed to ensure that the benefits of AI technologies serve the common good rather than private shareholders.

1. **Universal, equal and non-discriminatory access.** Public AI components – including compute, models, data, applications and software – should be accessible through open licenses, open standards or interoperability requirements, and embedded in open-source ecosystems.

2. **Mission-driven public goals.** Rather than prioritizing market competition for its own sake, public AI focuses on advancing public objectives and filling gaps left by market actors. Public AI infrastructure often generates digital public goods[10] – resources that create social value beyond purely market-based return.

3. **Public control.** To meet the standard of "publicness," AI infrastructure requires meaningful public oversight, public funding or direct public provision and operation. This may also involve citizen participation in governance or requirements for the public to hold majority stakes in successful innovations.

This framework aligns with the Paris Charter on AI in the Public Interest, signed at the 2025 AI Action Summit, which sets out the following policy vision:

> *"The benefits of AI in the public interest rely on building open public goods and infrastructure, providing an alternative to existing market concentration, ensuring democratic participation, enforcing accountability and developing environmentally sustainable solutions. To fulfil this vision, we focus on enabling conditions for infrastructure in areas with demonstrated benefits in the public interest."*[11]

10 Digital Public Goods Alliance, Digital Public Goods (2025), https://www.digitalpublicgoods.net/digital-public-goods

11 Elysee, The Paris Charter on Artificial Intelligence in the Public Interest, February 11, 2025, https://www.elysee.fr/en/emmanuel-macron/2025/02/11/the-paris-charter-on-artificial-intelligence-in-the-public-interest

The European Union must now move beyond high-level declarations and formalize public AI as a pillar of its digital strategy. This requires the European Commission to translate these commitments into a coherent regulatory and investment roadmap that prioritizes the delivery of public goods. To prove effective, however, a public AI strategy must be implemented across the different layers of AI infrastructure.

3 | Public AI as a full-stack policy

To be effective, a public AI strategy must move beyond piecemeal interventions and adopt an integrated view of the entire AI technology stack, rather than addressing individual layers in isolation. In the private sector, dominant firms consolidate market power through vertical integration across layers, shaping entire ecosystems in the process. The EU must adopt a similarly coordinated approach, embedding the core pillars of public AI – access, mission-driven goals and public control – across all levels of the stack. By doing so, Europe can build a resilient ecosystem that strengthens European sovereignty by reducing structural dependence on proprietary solutions. At its most basic level, the public AI infrastructure stack consists of five interconnected layers: compute, data, models, applications and software.

This stack primarily describes the supply side of AI: the infrastructure, tools and services that enable AI development and deployment. For effective AI policy, however, the demand side is equally important. Measures that increase AI adoption – including targeted support for public-sector use cases, incentives for small and medium-sized enterprises, standards and guidance for responsible deployment, and strategic public procurement – are essential to ensure that this infrastructure is used in ways that foster innovation, strengthen European sovereignty and advance public-interest goals.

To date, the European Commission and member states have largely focused public intervention on expanding computing capacity. These investments, however, should not be treated as an end in themselves, particularly since rapid rebalancing at this layer alone

Layer	Description	Public AI components
Compute	The foundational physical and software infrastructure that enables AI development, including processors, chips and data centers. Specialized processors or chips are combined with software frameworks and integrated into data centers. Computing power is provided either on site or through cloud platforms.	Public compute resources.
Data	The storage, processing and transfer of datasets used during training and post-training phases. This layer also includes real-time data inputs required at the inference stage to generate model outputs	Data commons for AI training.
Models	General-purpose or specialized AI models that function as reusable resources for a wide range of AI solutions. These models are typically deployed as cloud-based services.	Open-source AI models, open benchmarks.
Applications	User-facing systems and services in which AI models are embedded. Their operation requires additional computing capacity, often referred to as inference compute.	Public goal-oriented AI solutions, built on open-source AI models; open protocols for AI deployment.
Software	A cross-cutting layer spanning the entire stack, from proprietary optimization tools to open-source deep learning frameworks.	Critical software libraries shared as open-source software.

is unlikely given the concentration and capital intensity of compute markets. The European Union cannot achieve sovereignty through compute alone. There is therefore a risk that AI policy could reinforce existing dependencies by subsidizing proprietary AI solutions.

The AI Continent Action Plan, proposed in April 2025, represents an initial step toward a broader strategy that extends intervention to adjacent layers of the stack. This entails aligning measures on data access and quality, model availability and governance, support for critical software infrastructure, and oversight of how AI applications are procured and deployed, so that progress in one area reinforces advances in others.

Public AI policy should rest on the premise that, while a fully sovereign AI stack is neither achievable nor desirable, dependencies can be significantly reduced and AI systems can be more clearly oriented toward the public interest.

4 | The case for European public AI

The global context for AI development is increasingly framed as an"AI race." dynamic that has been shaping the global context for AI development. This dynamic is largely dominated by the United States and China, while the European Union is still defining its role. Unprecedented spending on computing power is driven by the scaling laws that underpin the dominant AI development paradigm, in which model performance improves as data volumes, compute capacity and model size increase. As a result, modern generative AI systems rely on highly data- and compute-intensive training methods.[12] Competition between the two leading powers, combined with these technological dynamics, is pushing both state and corporate actors toward ever larger infrastructure investments in pursuit of market power and geopolitical influence.

The window for intervention is rapidly closing. Astronomical investments are being made, control over key infrastructure layers is consolidating, and incumbents are erecting barriers to entry for new market players, including European firms. Without proactive measures, the benefits of AI technologies are likely to accrue to a privileged minority, while the risks – from disinformation to algorithmic bias – will disproportionately affect European citizens, communities and institutions.

The European Union needs to develop AI technologies and solutions, but on its own terms. Attempting to replicate the approaches of the United States or China is neither realistic nor desirable, even if the "AI arms race" defines the broader context. Europe lacks the resources to compete on those terms. Instead, it must articulate a third path for AI and build an alliance of like-minded partners around a shared, multi-

12 Jared Kaplan et al., "Scaling Laws for Neural Language Models," arXiv (2020), https://arxiv.org/abs/2001.08361

national development effort. This could include neighboring countries such as the United Kingdom and Norway, G7 partners such as Canada and South Korea, and emerging leaders including Brazil and India. Cooperation should involve the active pooling of resources – from compute infrastructure and datasets to software components – as well as joint development of open-source AI models. Such an alliance would allow mid-sized economies to achieve the scale required for frontier model development, turning sovereign AI into a shared, multinational asset.

European policymakers must also weigh the opportunity costs of large-scale AI investment. Every euro directed toward AI is a euro not invested in other policy priorities, particularly given that the assumed benefits of AI development and adoption remain unproven. Establishing a base layer of public digital infrastructure for compute access, data and model development should therefore be the cornerstone of the European Union's AI strategy.

The European Commission has acknowledged these challenges and responded with the **AI Continent Action Plan**, published in April 2025. The plan sets out the EU's ambition to position Europe as a "global leader in artificial intelligence," based on a distinct approach. While it frames AI development as a "race for leadership in AI," it also emphasizes the goal of fostering a "European brand of open innovation." Of the plan's five priority areas, three are particularly relevant to a public AI strategy: expanding Europe's computing capacity, improving access to high-quality data for AI use, and supporting AI deployment in strategic sectors. The remaining priorities focus on regulatory reform and skills and talent development. Financing is to be provided through the InvestAI Initiative, which aims to leverage €50 billion in public funding to mobilize an additional €150 billion from private investors and industry.

While the AI Continent Action Plan outlines key technical building blocks of a European AI stack – compute, data and models – it falls short of articulating a coherent public AI strategy. By treating these layers as discrete technical objectives rather than as components of an integrated public digital infrastructure, the Commission risks spending billions on a foundation that does not deliver greater competitiveness or strategic autonomy. Although the plan references European values and open-source collaboration, it lacks the mission-driven focus and explicit public orientation needed to ensure that these investments serve the common good.

The **Apply AI strategy**, introduced in October 2025, shifts attention to AI deployment across strategic sectors through 11 sectoral flagship programs. Absent a clear public AI mandate at its core, the planned investment of more than €1 billion risks functioning primarily as a subsidy for private firms. Without explicit requirements for universal access and public control over AI solutions, the proposed "AI first policy" is likely to accelerate the deployment of proprietary tools that fail to benefit the broader European social model.

A critical test of this strategic shift is the Frontier AI Initiative, which, from the perspective of public AI and European sovereignty, is the most consequential element of the broader approach. The initiative is premised on the view that "it is a priority to ensure that European models with cutting-edge capabilities reinforce sovereignty and competitiveness in a trustworthy and human centric manner." To meet this objective, open-source development must be established as a non-negotiable principle. Using AI factories and gigafactories to develop anything less than a family of permanently open, democratically governed "flagship models" would squander a rare opportunity to build independent AI infrastructure. Open-source principles are essential if such models are to serve as platforms for distributed innovation, including innovation in the public interest.

Finally, the **Data Union Strategy**, announced in November 2025, aims to expand access to high-quality data for AI development and address current shortages. It builds on earlier efforts, under which the Commission invested €336 million between 2021 and 2024 in 14 common European data spaces. These initiatives provide sector-specific frameworks for voluntary data-sharing.

Under the new strategy, the EU plans to launch a flagship initiative investing in secure, high-quality and in-

teroperable datasets that reflect European values and standards, enabling Europe to "use its own data" for AI development. Plans also include expanding the list of high-value datasets made available by the public sector, creating a federation of scientific data repositories under the European Open Science Cloud, and scaling linguistic datasets through the Alliance for Language Technologies (ALT-EDIC).

A critical evolution in this strategy is the introduction of data labs located within AI factories. These are intended to serve as the "institutional glue" linking data holders, AI developers and computing infrastructure. To align with a public AI framework, however, policymakers must ensure that these labs do not become mere service hubs for private startups. Instead, they should be grounded in the concept of a European data commons. The objective should extend beyond expanding access to include public control and a clear orientation toward AI applications that serve the public interest.

5 | The need for a Public AI intervention

The strategies announced by the European Commission provide an initial foundation for a European public AI approach. They correctly acknowledge the need for strategic intervention across the full AI stack. However, to effectively counter the growing concentration of power and ensure that AI development serves the public interest, these efforts require further strategic refinement and coordination. The following recommendations outline the necessary shifts in focus, governance and investment across computing power, AI models and data.

Compute

A central element of the AI Continent Action Plan is the expansion of computing capacity. As part of this effort, the European Commission is reinforcing the AI Factories initiative, which upgrades Europe's existing network of supercomputers with processors optimized for artificial intelligence. These facilities are intended to make advanced compute resources available to small and medium-sized enterprises, researchers and startups, enabling them to build, train and test AI systems that would otherwise be out of reach. By late 2025, 19 AI factories had been selected across Europe – seven announced in December 2024, six in March 2025 and a further six following a third cutoff in mid-2025.

Alongside this initiative, the Commission has announced plans for AI gigafactories: very large data centers dedicated to developing and training complex AI systems at a scale not previously available in Europe. Each gigafactory is expected to be equipped with more than 100,000 high-performance AI processors, comparable to H100-class GPUs, compared with up to 25,000 processors planned for the largest AI factories. Up to five gigafactories are envisioned, supported by a dedicated €20 billion fund. This funding is part of InvestAI, a broader €200 billion investment program for artificial intelligence, and is expected to cover roughly one-third of the capital expenditure for each facility.[13]

While AI factories and gigafactories share a common technological foundation, gigafactories represent a qualitative shift in both scale and purpose. According to the European Commission, they are explicitly designed to enable Europe to train and deploy AI models "with hundreds of trillions of parameters."

By contrast, the Apply AI strategy extends the role of gigafactories beyond frontier model training and deployment. It frames them as a compute backbone for a wider range of sectoral applications, including training models for defense and space, and accelerating the development of models and software platforms for autonomous driving and vehicle management systems. In this interpretation, gigafactories function as shared compute infrastructure serving diverse use cases, including public-interest AI applications, rather than focusing exclusively on state-of-the-art foundation models.

13 European Commission, The AI Continent Action Plan (Brussels: European Commission, 2025), https://digital-strategy.ec.europa.eu/en/library/ai-continent-action-plan

At present, however, the two strategies are not fully aligned on the purpose of AI gigafactories. The AI Continent Action Plan implicitly assumes an anchor-customer model centered on one or a small number of labs with sustained, large-scale training workloads. Europe's current demand profile makes this assumption difficult to sustain. Globally, only a handful of labs – including OpenAI, Google, Anthropic, xAI and Meta – consistently generate such workloads. In Europe, there is currently only one leading frontier lab, Mistral, capable of playing a comparable role.[14] These companies have acted as anchor customers in the existing global compute build-out, ensuring high utilization rates. In Europe, by contrast, a multi-client model that aggregates demand from a diverse set of users – including companies, startups, academia and the public sector – is likely to be better suited to AI gigafactories.[15]

A multi-client approach serving low-to-moderate AI workloads also aligns more closely with public AI needs. Many public-interest AI applications do not require extreme levels of compute. Moreover, developing strong open models that can serve as a foundation for public AI applications does not demand the same compute intensity as leading U.S. frontier models, such as GPT-5, which has been reported to require around 100,000 GPUs across two training runs.[16] Taken together, this suggests that EU investments in compute should be more clearly oriented toward public AI goals and that gigafactories should adopt an operating model aligned with Europe's existing and foreseeable demand.

Policy recommendations

Strengthen the focus on public AI goals. Investments in computing power – and in particular the gigafactories initiative – require a much clearer orientation toward public AI objectives. While AI factories and gigafactories should support a broad range of AI development efforts, the strategic importance of public AI must be explicitly recognized and prioritized. The absence of a mission-driven focus on using compute and data investments to enable state-of-the-art European public AI applications is the most significant shortcoming of the AI Continent Action Plan. The communication also fails to recognize the value of open-source AI, despite provisions in the AI Act that create favorable conditions for such systems.

Adopt a clear operating model for AI gigafactories. Given the absence of single, large anchor customers in Europe, a multi-client model that aggregates demand from a diverse set of users appears more appropriate. To serve this user base, gigafactories should focus particularly on commercial and public AI applications over purely research-driven use cases and operate as one-stop shops. This implies offering more than raw compute, including permanent storage for digital commons such as open datasets and models, public inference services, the ability to run commercial and operational workloads on site, and a basic layer of onboarding, curated software stacks and technical support. The European Commission should explicitly consider such an operating model when evaluating gigafactory proposals.

Models

Demand-driven development of AI models – both general-purpose and specialized – should be at the core of the European AI strategy. A stronger focus on model development would help capture the value generated by public investment in compute and data and more importantly, channel these resources into building viable alternatives to dominant commercial AI systems.

14 Julia Christina Hess and Felix Sieker, Built for Purpose? Demand-Led Scenarios for Europe's AI Gigafactories (policy brief, Bertelsmann Stiftung and interface e.V., October 22, 2025), https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/Final_251021-ai-giga-factories.pdf

15 Julia Christina Hess and Felix Sieker, Built for Purpose? Demand-Led Scenarios for Europe's AI Gigafactories (policy brief, Bertelsmann Stiftung and interface e.V., October 22, 2025), https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/Final_251021-ai-giga-factories.pdf

16 Matthew Griffin, "OpenAI GPT-5 Is Costing $500 Million per Training Run and Still Failing," Fanatical Futurist, May 2025, https://www.fanaticalfuturist.com/2025/05/openai-gpt-5-is-costing-500-million-per-training-run-and-still-failing/

To date, European Commission support for model development has been limited. The Large AI Grand Challenge, launched in November 2024, awarded four companies €1 million and eight million computing hours in mid-2025. Awardees were given 12 months either to release their models under "an open-source license for non-commercial use," or to publish their research findings. OpenEuroLLM, a three-year project launched in early 2025, received nearly €40 million to develop a family of high-performance, multilingual, open-source large language models. At the national level, a range of public and public-private initiatives have trained open-source or open-weights models, including Spain's Alia and Aina, Poland's PLLuM and Bielik, Sweden's GPT-SW3 and Germany's Teuken. These efforts represent important first steps, but on their own they are insufficient to produce models that are both capable and globally competitive.

The Frontier AI Initiative, announced as part of the Apply AI strategy, is intended to "accelerate progress" in frontier AI capabilities. The strategy also points to the development of sector-specific frontier models, including for space, manufacturing and Earth systems. At present, however, there is limited clarity on how the initiative will build on existing efforts, and coordination across programs remains a significant challenge.

The Initiative should explicitly prioritize the development of a family of European public foundation models: permanently open-source, democratically governed models that aspire to remain at or near the frontier of AI capabilities. As a general-purpose technology, such models could become the engine of public value generation and provide the technological backbone for a European generative AI ecosystem.[17]

However, the European AI strategy need not choose between large foundation models and smaller alternatives. Both AI development paths are viable and mutually reinforcing. Foundation models, referred to as general-purpose AI, sit at the core of the current AI paradigm. Smaller models, by contrast, require less computational power and offer advantages in efficiency, accessibility and customization, making them well suited for deployment on resource-constrained devices, in edge computing environments and in domain-specific applications.

This two-track approach fosters an open-source AI ecosystem that benefits from a central, capable model while also investing in more specialized, sustainable solutions. It also combines the general-purpose capacity of a central flagship model with more specialized, sustainable solutions tailored to specific needs. Given the rapid pace of change in AI development, public AI policy must be grounded in a clear understanding of technological advances that enable more affordable yet capable models, alongside targeted investments that strengthen the entire open-source AI ecosystem and support innovation in the public interest.

Policy recommendations

Ensure open-sourcing and model transparency. The Frontier AI Initiative needs to commit to open-sourcing models and ensuring full transparency, in compliance with the AI Act. Models should be made available under terms that grant users the freedom to use, study, modify and share them. This requires that model code, architectures and parameters be openly accessible. With respect to data, open models should rely on open datasets wherever possible, accompanied by comprehensive documentation on data sources and composition. Open-sourcing is also a pro-competition measure: publicly funded AI technologies can be used by European businesses, public institutions and communities without creating new dependencies that put European sovereignty at risk.

Prioritize the development of a family of European public foundation models. These flagship models need to be state of the art in terms of their capability. While matching the strongest commercial models may not be feasible – or even desirable – recent developments suggest that first-mover advantages are diminishing. Models such as DeepSeek and K2-Think illustrate how techniques like model distillation can

17 Felix Sieker et al., "Public AI - white paper", Bertelsmann Stiftung (2025), https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/public-ai-white-paper-a-public-alternative-to-private-ai-dominance

enable the development of capable systems with relatively limited resources.[18] The initiative should prioritize the development of a comprehensive suite of language technology solutions including voice models, multilingual translation systems and BERT-style encoders. These should be complemented by specialized models for other scientific domains, such as the Earth system model referenced in the Apply AI strategy.

As a first step, the initiative should define a clear model development pathway that aligns available infrastructure and computing capacity with specific AI development objectives. In parallel, targeted research into alternative AI development approaches should be supported, with a focus on improving training efficiency. Such innovation could lower energy demands and operating costs, making advanced AI capabilities more accessible to public institutions and research organizations.

Invest in software commons and open standards. Targeted investment is needed to support the development of key software components for AI development as digital public goods. Despite their central role, many widely used tools – from Python libraries for data preparation such as pandas to machine-learning libraries like scikit-learn – face persistent sustainability challenges. Public funding can play a key role in ensuring their long-term maintenance, security and continued development. Previous examples include France's €32 million commitment to support scikit-learn under its national AI strategy and the German Sovereign Tech Agency's backing of core Python libraries.

The development of open standards and benchmarks is an equally important enabling measure. Open standards are essential for orchestrating models, including by ensuring interoperability between various proprietary and nonproprietary solutions. And open benchmarks are essential for measuring model capabilities and societal impact. While existing benchmarks largely emphasize technical capabilities, future efforts should also evaluate how well AI systems advance public objectives, particularly in regulated sectors where the stakes for public safety or general interest are high.

Data

Major AI developers make little effort to share high-quality, reusable datasets. This holds even for companies that promote a degree of openness by releasing open-weights models, as in the case of Mistral. Instead, firms seek competitive advantage through proprietary data sources, including user-generated and personal data from social networks and online platforms, as well as data secured through exclusive commercial agreements. Public web content remains an important data source, but there are signs that the social contract underpinning the open web is weakening, as content owners increasingly withdraw consent for their material to be crawled.

These dynamics contribute to a negative feedback loop, described by Stefaan Verhulst as a "data winter" - declining willingness to treat data as a common good. The effect is particularly damaging for open-source AI development, which lacks access to proprietary data. A fundamental asymmetry has emerged: dominant non-EU actors operate with far fewer constraints, often collecting data in ways that disregard privacy or copyright standards that public-interest actors in the EU cannot, and should not, emulate. The result is a widening gap in which European labs are increasingly starved of training data.

To move beyond these limitations, the Data Union Strategy must address the gap that currently handicaps European developers. Competitive AI systems cannot be built on curated sector-specific datasets alone, and the European common data spaces have not yet provided a sufficient solution. Addressing this challenge also requires greater legal clarity on the use of publicly available data by EU-based developers. A public AI data strategy should therefore follow a two-track approach: establishing legal certainty for the use of public data, while building a data commons of high-quality, specialized datasets that complement public web resources.

18 Frank Nagle and Daniel Yue, "The Latent Role of Open Models in the AI Economy," SSRN (2025), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5767103

This means data-sharing efforts must shift away from simply increasing data volumes and toward improving data quality and establishing governance mechanisms that enable equitable and sustainable sharing while preventing extractive use. The objective should be to support a public AI life cycle in which sharing high-quality datasets underpins the development of open-source AI models.

In recent years, there have been few public interventions aimed at enabling the sharing of high-quality data. At the same time, the European common data spaces – a flagship initiative of the previous European Data Strategy – were not designed to provision datasets or guarantee access to public data. One notable exception is the Alliance for Language Technologies (ALT-EDIC), launched in March 2025. Its goal is to expand the availability of European language data and to develop a shared infrastructure for language resources, supported by €4 million secured by the EDIC. A further €88 million from the Digital Europe program has been allocated to four ALT-EDIC projects, including OpenEuroLLM and LLMs4EU, both focused on model development. High Performance Language Technologies (HPLT), a Horizon Europe project with a budget of €4 million, also contributes by developing open language datasets.

The recently announced Data Union Strategy signals a new emphasis on data-sharing for AI development. However, it still lacks mechanisms that would enable European AI developers to use publicly available data at scale. And the proposed data labs require a stronger public orientation. They should work closely with the public sector on both the supply and demand sides of data, prioritize the needs of open-source AI development – particularly under the Frontier AI Initiative – and champion the concept of a European data commons, treating data as a public good.

Policy recommendations

Legal clarity on AI model training. Europe needs to leverage flexibility already embedded in its legal framework for training AI models on publicly available data. The existing legal bases for text and data mining, set out in Articles 3 and 4 of the DSM Directive (2019/790), should be explicitly confirmed as covering AI model training on lawfully accessible data and lawfully derived datasets. In addition, the European Commission should clarify that the sharing of derived datasets prepared for AI training is permissible for beneficiaries under Article 3, including for purposes of collaborative research and reproducibility. Such clarification would support greater harmonization across AI development efforts in Europe, improve the efficiency of cross-border research collaboration and help ensure the economic impact of publicly funded AI development.

Public data commons. Data labs should be tasked with stewarding a European data commons built around three governance principles: enabling access through clear data-sharing frameworks and permission interfaces; ensuring collective governance through trusted institutions and in partnership with relevant communities; and generating public value through public-interest-oriented licensing models.

Data commons institutions would play a gatekeeping role, supporting various types of data and applying flexible governance models, ranging from open data and open access to more restricted arrangements that protect individual and collective data rights. A European data commons should also provide the infrastructure needed to host AI training datasets and support public sector institutions in this regard. Such an initiative should unlock existing datasets, focusing on those of high quality and relevance. It should also create new datasets for priority use cases where data is currently unavailable. This includes not only data for pretraining, but also instruction and evaluation datasets. Achieving this will require mechanisms to identify high-relevance use cases in collaboration with the public, supporting a demand-driven approach to AI development. The approach must shift from simply releasing as much data as possible to intentionally creating high-quality, purpose-built datasets for AI training.

6 | Conclusion: Public AI and purposeful deployment of AI technologies

European investment in digital technologies has often favored industry-led, technology-driven projects, with less emphasis on problem-oriented approaches and concrete societal needs. By adopting a more discerning and realistic view of emerging technologies, EU funding could prioritize initiatives that respond to genuine needs and deliver tangible value. This requires a careful balance between supporting industry and advancing public value objectives.

There is a risk that a European AI deployment strategy will embrace an uncritical "AI first" principle, accelerating the rollout of technologies without a clear sense of purpose. A public AI strategy offers a different path: one centered on purposeful, demand-driven deployment, grounded in clearly identified needs.

Purposeful deployment needs to be a core principle of a public AI strategy. It ensures that AI is adopted only where there is clear evidence of benefit and alignment with public interest goals. Applied AI solutions need to address real societal challenges and serve communities across Europe. To this end, public funding for digital technologies should draw more explicitly on mission-oriented research and innovation policy and allocate sufficient resources to addressing defined problems, rather than pursuing technological ambition for its own sake.

A successful European public AI strategy will create strong governance mechanisms capable of creating precise and purposeful deployment roadmaps. This includes the willingness to identify cases in which AI is not an appropriate solution and to concentrate deployment efforts where they demonstrably create value.

Commissioned by

January 2026
Publisher | Bertelsmann Stiftung
Carl-Bertelsmann-Straße 256 | 33311 Gütersloh
Phone +49 5241 81-0 | www.bertelsmann-stiftung.de
Responsible | Dr. Felix Sieker | Project Manager
Digitalization and the Common Good
Phone +49 30 275788-156
felix.sieker@bertelsmann-stiftung.de
Layout | Nicole Meyerholz, Bielefeld
Editing | Barbara Serfozo, Berlin
Cover illustration | Jakub Koźniewski, Warsaw

DOI 10.11586/2026001