

Maria Nowicka
September 2026

Powering AI

*How Europe can align its AI data centre
buildout with energy realities*

Imprint

Published by

Friedrich-Ebert-Stiftung e. V.
Godesberger Allee 149
53175 Bonn
Germany
info@fes.de

Issuing Department

Friedrich-Ebert-Stiftung | Competence Centre Climate and Social Justice
Cours Saint Michel 30e | 1040 Brussels | Belgium

Responsibility for this Publication in the FES

Richard Probst, Director of FES Competence Centre Climate and Social Justice

Editor

Reghina Dimitrisina, Policy Advisor,
FES Competence Centre Climate and Social Justice
reghina.dimitrisina@fes.de

Contact

justclimate@fes.de

Design/Layout

pertext | corporate publishing
www.pertext.de

Cover picture

AI generated image

The views expressed in this publication are not necessarily those of the Friedrich-Ebert-Stiftung (FES). Commercial use of media published by the FES is not permitted without the written consent of the FES. Publications by the FES may not be used for electioneering purposes.

September 2026

© Friedrich-Ebert-Stiftung e. V.

ISBN 978-3-98628-906-5

Further publications of the Friedrich-Ebert-Stiftung can be found here:

➤ www.fes.de/publikationen

Maria Nowicka
September 2026

Powering AI

*How Europe can align its AI data centre
buildout with energy realities*

Contents

- Context: Where Strategic Interests Clash –
AI, Global Power and Energy Transition 3
- Introduction 4
- AI Clusters as New Industrial Actors 6
- Mapping Evolving Energy Bottlenecks for AI Infrastructure 8
- Policy Recommendations 11
- Conclusion 13
- Bibliography 14

Abstract

This policy brief examines how the rapid build-out of AI compute clusters is reshaping Europe's electricity systems and what this means for the EU's broader tech-sovereignty agenda. AI-optimised data centres differ from traditional facilities both in hardware and in how workloads are scheduled: they concentrate very high-density accelerator hardware, are increasingly dominated by latency-sensitive inference, and are often sited close to metropolitan demand hubs. As a result, they behave as a new class of industrial-scale load that requires large, firm capacity at specific grid nodes, with limited short-term flexibility unless training and batch jobs are actively managed around system conditions.

The brief identifies three emerging bottlenecks. Firstly, ambitious AI infrastructure plans risk creating idle capacity and stranded grid investments if utilisation trajectories remain opaque and volatile. Secondly, without deliberate incentives and baseline obligations, the flexibility potential of schedulable AI workloads is likely to remain underused, even though it could help relieve congestion and absorb surplus renewables. Thirdly, current siting and permitting approaches, including proposed data-centre acceleration zones, still treat AI clusters much like conventional IT capacity and pay insufficient attention to constrained grids, local resource limits, and competing demands for clean power.

Against this backdrop, the paper argues that AI clusters must be recognised and regulated as distinct industrial actors on the grid. It proposes three directions for EU policy: explicitly defining AI clusters as a separate category in relevant frameworks; linking public support and fast-track treatment to transparent utilisation and verifiable flexibility commitments; and aligning siting decisions for AI-specific capacity with grid development plans and low-carbon generation. These choices will determine whether Europe's AI sovereignty agenda reinforces or undermines its path towards climate neutrality.

Context: Where Strategic Interests Clash – AI, Global Power and Energy Transition

Over the past decade, the scale and speed of AI deployment have transformed the technology from a niche research topic into a structuring factor of global power. Breakthroughs in large language models, both in training and in large-scale inference, have gone hand in hand with a rapid build-out of AI compute infrastructure. As a result, control over the AI value chain – from chips and cloud platforms to data and energy-hungry AI clusters – is now front and centre for both industrial competitiveness as well as national and economic security.

In the United States, frontier AI development is increasingly concentrated in a small group of frontier AI labs, many of which also operate as hyperscale cloud providers that control a large share of global compute capacity and leading foundation models. Recent data show that of the roughly 23 gigawatts of new data centre capacity under construction worldwide, about three quarters was located in the U.S., underscoring how strongly the current wave of AI infrastructure buildout is skewed towards the American market (Bloomberg, 2026). A growing share of this capacity is being built ‘behind the meter’, with operators procuring dedicated onsite generation rather than relying solely on the public grid. Recent industry analyses suggest that U.S. data centres are on track to add tens of gigawatts of behind-the-meter generation capacity by the end of this decade, including gas fired plants and other local sources, with some facilities explicitly aiming to operate largely off-grid to sidestep interconnection bottlenecks (SemiAnalysis, 2026). In parallel, emerging White House policies treat advanced AI capabilities as strategic assets, signalling that access to them is now seen as part of national security considerations.

China, for its part, has launched an explicit drive to expand domestic computing power and reduce dependence on foreign technology. Unlike U.S. providers, however, Chinese operators currently have no access to the most advanced logic and memory chips due to export controls and restrictions on extreme ultraviolet lithography and must rely on domestically produced or legacy-node accelerators. China’s national plans aim to increase aggregate computing capacity by more than a third in just a few years and to connect regional “intelligent computing centres” into a unified nationwide network. Central and provincial authorities support AI-optimised data centres based on domestic chips through subsidies and siting incentives, encouraging operators to “build ahead of demand” and treat AI compute as a long-term strategic infrastructure. While this has already

produced pockets of overcapacity and under-utilised clusters in some regions, it underlines Beijing’s intent to anchor future economic growth and state power in large-scale, domestically controlled AI infrastructure.

Against this backdrop, the European Union is recalibrating its own position. Its recently proposed European Tech Sovereignty Package brings together distinct but complementary pillars. The Chips Act revision targets the supply and demand for advanced semiconductors in Europe. The Cloud and AI Development Act proposal governs demand for those chips and the emergence of “sovereign” cloud and AI infrastructure, through instruments such as AI Factories and AI Gigafactories and new criteria for “sovereign” services. The EU Open Source Strategy addresses the software layer, seeking to grow open components and ecosystems on top of this infrastructure, while the Strategic Roadmap for Digitalisation and AI in the Energy Sector focuses on sustainability and energy system integration by linking data centres and AI clusters to decarbonised generation and grid planning. Together, these instruments connect hardware, infrastructure, software and energy into a tentative but more coherent tech sovereignty framework. At the same time, ongoing debates on sovereignty requirements for cloud services and the need to expand European-controlled data centre and AI compute capacity show that the EU is seeking to reduce structural dependence on a small number of non-EU providers for its critical workloads.

Recent U.S. restrictions on foreign access to frontier models have reinforced this drive for sovereignty. By demonstrating that access to cutting-edge AI systems can be curtailed overnight for non-nationals, these measures turn reliance on external compute and model providers into a strategic vulnerability. In Europe, sovereignty over AI compute and models is therefore inseparable from sovereignty over the energy and grid capacity that make them possible: on already constrained electricity systems, decisions about what kind of AI capacity to host, where to site it, and how it generates added value become strategic choices rather than neutral technical details.

Introduction

As AI moves into large-scale deployment, the buildout of AI optimised data centres is becoming one of the central drivers of electricity demand, grid planning and decarbonisation pathways in Europe. This brief focuses on the physical layer of the AI boom: how specialised AI clusters change power demand patterns and what this implies for European energy policy.

The scale of the shift is already visible in global projections. Recent assessments by the IEA indicate that global sectoral electricity demand from AI grew by around 17% last year only, and that global data centre demand could roughly double by 2030, with AI as a major driver of this growth. In its scenarios, total data centre electricity use reaches around 945 TWh by 2030, a level comparable to the current annual consumption of Japan (IEA, 2026). Industry and academic analyses converge on the expectation that AI clusters, built around high-density AI accelerators and liquid cooling, will grow much faster than traditional IT capacity. Some projections imply a tripling of AI-related electricity demand by the end of the decade (IEA, 2026). These trajectories suggest that energy is not just a constraint but the decisive factor for whether AI strategies are feasible, affordable and compatible with net-zero commitments, and thus determining whether Europe's vision of tech sovereignty can be realised in practice.

First environmental and system implications are emerging as well. The IEA warns that global emissions from data centres are on track to roughly double by 2030 if current trends continue, even under scenarios with significant decarbonisation of power supply. In Europe, analysts highlight that, without strict rules tying new data centre demand to additional renewables and grid upgrades, this demand can lock in fossil gas or displace renewables that other sectors would have used, thereby raising emissions. Combined with the fact that about a third of EU's total electricity generation still powered by fossil fuels and with existing bottlenecks around new grid connections across Member States, Europe's ambitious plans for AI compute expansion are likely to hinge on policymakers' ability to effectively connect the dots between technological sovereignty and energy policies (Eurostat, 2026).

Meanwhile, regions that were initially seen as global pioneers in AI and data-centre buildout are now facing the consequences of expanding AI infrastructure without specialised energy planning. In China, recent analyses

document overbuilt and underutilised AI computing capacity in some provinces, as AI clusters were rolled out faster than actual workloads and integration into national power and data markets. In the United States, the latest reports note that up to 13% of data-centre electricity consumption is associated with unused infrastructure, underscoring the risk of idle capacity (IDCA, 2026). The North American Electric Reliability Corporation has issued a Level-3 alert on "large load" challenges, explicitly warning that rapid data-centre growth in certain regions poses reliability risks if not matched by grid reinforcement and new capacity. The U.S. recently saw its first statewide data centre moratorium in New York and a grid access pause in Texas, with local communities across the country engaging in citizen-led initiatives such as the Brockovich Data Centre campaign mapping local impacts on water resources, pollution, and grid stress.

In this context, the EU's Tech Sovereignty Package and related initiatives such as the AI Continent Action Plan and AI Gigafactories (AIGFs) – with the latter now advancing into procurement and the selected consortia expected to be announced by November 2026 – have put large-scale compute clusters at the heart of Europe's digital and industrial strategies. Yet these initiatives have so far said little about how much additional electricity demand they imply, how quickly grids can accommodate it, or how AI-specific clusters should be integrated into already constrained systems.

These tensions are beginning to surface explicitly in EU-level debates. Brussels policymakers increasingly acknowledge that rapid data centre growth must be squared with Europe's decarbonisation agenda. As Energy Commissioner Dan Jørgensen put it *"we very much expect data centres (...) to support our Green Deal in Europe and transition towards more clean energy, both nuclear and renewables"*, signalling that the planned compute capacity is politically framed as needing to contribute to, rather than undermine, the green transition goals. The Strategic roadmap proposed as part of the Tech Sovereignty Package aims to address this nexus but stopped short of addressing specific challenges related to system integration of AI-specific clusters. At the same time, projections indicate that the planned AI infrastructure pipeline in Europe implies total data-centre capacity that is roughly 2.3 times higher than the current installed capacity, making electricity supply and grid integration a decisive factor for the feasibility of the strategy (IEA, 2026).

Against this backdrop, this policy paper examines why AI compute clusters, understood as specialised data-centre facilities that host large-scale AI workloads, require specific conditions from the perspective of energy integration, and what kinds of bottlenecks need to be addressed. It explains how AI clusters differ from traditional data centres, maps the emerging energy challenges linked to their expansion, and puts forward policy recommendations that can be embedded in ongoing EU initiatives, including initiatives within the EU Tech Sovereignty Package, as well as the Data Centre Efficiency Package. Where relevant, this brief also assesses the extent to which the AI Gigafactory tender criteria published in July 2026 address these bottlenecks. The AI Gigafactories initiative is particularly relevant, as it constitutes the EU's flagship programme for deploying large-scale AI compute clusters – the facilities at the very centre of this analysis. The goal of this brief is to inform policymakers and highlight the specific energy considerations that should shape Europe's AI infrastructure strategies.

Methodology

This brief draws on a targeted review of recent policy documents, technical studies, industry analyses and public data on AI compute infrastructure, electricity demand and data-centre development in Europe, the United States and China. It combines evidence from EU policy files, international energy and technology assessments, and emerging literature on AI workloads, infrastructure bottlenecks and grid integration. In addition, the analysis was informed by expert consultations conducted during the drafting process with stakeholders from academia, think tanks, civil society and public-policy institutions, including Felix Sieker (Bertelsmann Foundation), Jan Peter Kleinhans (OECD), Gauthier Roussilhe (Sciences Po) and Nicoleta Kyosovska (CEPS).

AI Clusters as New Industrial Actors

Over the past few years, most European policy discussions on data centres have focused on their aggregate energy use and efficiency, treating them largely as a specialised form of commercial infrastructure. The recent AI boom has fundamentally changed this picture. AI clusters are emerging as a new class of industrial scale demand, defined by both how they are built (hardware) and how they are used (workloads), in ways that distinguish them from the traditional data centres that informed earlier policy.

On the hardware side, traditional data centres were dominated by servers centered around CPUs (*Central Processing Units*) in air-cooled halls, with typical rack power densities

of 5–10 kW and a significant share of total energy going into building services. AI clusters, by contrast, are built around dense concentrations of GPUs (*Graphic Processing Units* – such as NVIDIA’s H100) and other components that push power densities beyond 120 kW per rack (Semi-Analysis, 2026). Because of this shift, AI-oriented data centres are typically designed so that IT equipment accounts for more than 60 per cent of total power, and the combination of highly concentrated IT load with comparatively lean cooling and power overhead makes any disturbance in these subsystems much more likely to affect compute performance than in many traditional data centres (X. Chen et. al, 2025). The table below summarises

Comparison of architectures – AI compute cluster and traditional data centre

Table 1

	AI compute cluster	Traditional data centre
Compute	Hardware: Specialised AI accelerators (GPUs/TPUs) with HBM Parallelism: Deep, large homogenous tasks Workload: Massive parallel tasks	Hardware: General-purpose CPUs and minimal accelerators Parallelism: Many small independent tasks Workload: Mixed, latency-sensitive applications
Memory and storage	Hierarchy: HBM-local SSD-central HDD/object storage Priority: Keep accelerators fed; data proximity is key Access pattern: Sequential, large-batch reads	Hierarchy: Server DRAM-thin SSD layer-central HDD Priority: Cost optimisation and request throughput Access pattern: Random, small, frequent requests
Networking	Topology: Dense east-west compute fabric Interconnects: NVLink, InfiniBand, and high-speed Ethernet Scale: Synchronises thousands of accelerators	Topology: North-south user to server flows Interconnects: Conventional Ethernet Scale: Fewer high-speed ports per rack
Power delivery	Density: 3–10 times the conventional rack power Distribution: HVDC + advanced power electronics Design driver: Managing conversion losses at scale	Density: Standard rack power levels Distribution: Conventional AC-centric chain Design driver: Smooth predictable loads
Cooling	Method: Liquid (direct-to-chip or immersion) Threshold: Design assumption higher than 30 kW/rack	Method: Air cooling (room level or in-row) Threshold: Adequate up to 15–30 kW/rack

The original analysis featuring a full breakdown of hardware architecture of AI compute clusters can be found here: <https://www.interface-eu.org/publications/ai-compute-energy-bottlenecks#chapter-1-how-ai-hardware-uses-energy-inside-large-ai-compute-clusters>

the main physical characteristics of traditional data centres and AI clusters side by side.

On the workload side, the key distinction is between training (development) and inference (deployment), which captures how the installed hardware is actually used. AI training runs high-intensity computation processes for hours or days; it is energy hungry but often allows some flexibility in timing and location, as long as training deadlines and service-level objectives are respected. Inference applies trained models to live user interactions – such as chatbots or search – and therefore creates a near-constant, latency-sensitive draw with very limited short-term flexibility. Because AI operators typically prioritise inference to meet service-level agreements and fill remaining capacity with training jobs, the least flexible component of AI demand is also the one that tends to grow fastest.

In general terms, AI clusters in Europe are likely to fall into three broad categories: those dominated by training workloads, mixed clusters that combine training with large-scale inference for commercial services, and inference-only clusters focused on serving latency-sensitive applications.

In practice, clusters use scheduling and orchestration systems to allocate AI workloads. Recent studies note that operators generally prioritise online inference to satisfy service level agreements, with training and batch workloads filling the remaining capacity (F. Acun et. al, 2026). Broader AI workload analyses, including data aggregated by Epoch AI, indicate that as models moved from experimentation into widely deployed products with a strong increase in users globally, a growing share of global AI compute is devoted to inference rather than training.

From an energy-system point of view, this workload distinction matters because it shapes the electricity consumption patterns, siting choices and flexibility potential of AI clusters (M. Nowicka, 2026). Training-heavy AI clusters can, in principle, be located closer to cheap, abundant low-carbon generation, since they are less constrained by latency and proximity to end users. Inference-heavy clusters tend to gravitate towards metropolitan regions and network hubs, where they can minimise response times and integrate closely with existing cloud and content-delivery infrastructure.

The combination of a highly complex and power dense hardware setup, inference-dominated workloads and metropolitan siting gives AI clusters a distinctive profile. They demand large amounts of firm capacity at specific grid nodes, maintain high power draw around the clock and have limited short-term flexibility unless training and batch workloads are deliberately scheduled around system conditions. The following subsection describes this pattern in terms of scale, concentration and temporal profile, laying the ground for understanding why AI clusters need to be treated differently from traditional data centres in energy policy.

Mapping Evolving Energy Bottlenecks for AI Infrastructure

Europe's AI infrastructure plans are taking shape at a time when electricity systems are already under stress from delayed grid investments, uneven renewables integration and lingering fossil dependencies. Recent analyses show that AI clusters reinforce these pressures at three levels: generation, networks and governance. Even without precise forecasts of how much electricity AI clusters in Europe will ultimately consume, experience from early adopter regions and emerging European evidence already reveal three types of bottlenecks: constrained grid access and long connection lead times, risks of underutilised reserved capacity, and unresolved questions around flexibility, siting and infrastructure planning.

Idle Capacity and Stranded Assets

The first bottleneck is the risk that Europe builds AI-optimised capacity and supporting grid infrastructure that fails to reach high utilisation over its lifetime. AI workloads themselves make this outcome more likely. Training workloads require steady and high power, but they arrive in waves linked to model development cycles and can fall sharply when architectures or priorities change. Inference workloads, on the other hand, are bursty in their power draw but are characterised by 'always-on' power draw since they stem from serving live user queries. Together, these characteristics mean that AI clusters have a structurally higher risk of demand volatility and partial utilisation than many traditional industrial loads.

In Europe, this risk intersects with highly ambitious infrastructure plans. The AI Continent Action Plan and the subsequent CADA proposal envisage tripling European data-centre capacity within five to seven years, with a strong emphasis on AI compute cluster expansion. Yet most of the pipeline consists of project announcements and site plans rather than firm, long-term workload commitments. Crucially, emerging evidence of underutilised capacity in AI clusters has not slowed the pipeline: in both the U.S. and China, announced AI data-centre projects continue to grow at scale, with little systematic scrutiny of whether the resulting infrastructure will be fully used (A. Kaul, 2026).

The AI Gigafactory tender illustrates the pattern at a very concrete scale: each selected consortium is contractually

bound to install 75,000–100,000 H100 equivalent AI accelerators and scale up to 120–150 MW of IT load, deploying the bulk of that capacity within an 18-month set-up period, followed by a further 12-month expansion window. To put this into perspective, 75,000–100,000 AI accelerators represent a compute installation 3–4 times larger than today's leading European supercomputers. A 150 MW IT load alone is comparable to the continuous electricity demand of roughly 400,000 average EU households, yet the AIGF tender includes no requirement to disclose or even track the expected utilisation trajectories, anchor-customer commitments, or workload pipelines to justify this scale.

The CADA proposal, on the other hand, partially recognises that utilisation matters. In its list of grand challenges for the new Cloud and AI Leadership Initiative, the proposal recognises the need to improve utilisation and efficiency of cloud and data-centre resources, signalling awareness that idle capacity undermines both competitiveness and sustainability. However, this recognition remains generic. CADA does not spell out how utilisation should be measured for AI-specific workloads, nor does it distinguish between the utilisation patterns of AI clusters and those of traditional IT data centres. As technical analyses show, AI workload utilisation is shaped by factors – such as training cycles, inference latency constraints and rapid hardware turnover – that differ significantly from typical enterprise or cloud workloads. Without reflecting these specificities, policy debates on utilisation risk being too general to result in concrete criteria for AI infrastructure planning.

Missed Flexibility Potential

This utilisation problem is closely linked to flexibility. The same workload characteristics that make AI clusters prone to underutilisation also mean they could, in theory, provide meaningful demand response services. Training workloads are long running but often schedulable; they can be paused, slowed or shifted in time and, sometimes, in location, within certain bounds. Inference, by contrast, is latency sensitive and behaves like an "always on" industrial load, with tight limits on short-term curtailment. For training-heavy clusters, this raises the question of whether underutilised capacity could be repurposed for other workloads or shared across users. For mixed and inference-heavy clusters,

the key issue is how much of the training and batch component can realistically be exposed to demand-response programmes.

Recent technical work confirms that, in real demand-response programmes, AI clusters can reduce power draw from training and other non-latency-critical workloads by around 20–50% over short intervals without violating service-level agreements (F. Acun et al., 2026). First projects with large operators show that shifting or briefly reducing AI workloads during grid stress events is technically feasible and can be done at scale (T. Mann, 2025). From a system perspective, this means that even if average utilisation of AI clusters remains below the standard 35–45% observed in traditional data centres, the “slack” capacity could be turned into a resource for balancing and congestion management, if appropriate incentives and rules are in place.

At EU level, the Strategic Roadmap recognises data centres as ‘good candidates’ for flexible connection agreements and participation in flexibility markets, but it leaves the design of such agreements to Member States and does not define AI-specific obligations or safeguards. Deliberate AI workloads scheduling rules, centred on system conditions and aligned with periods of high renewable output or curtailed during scarcity events, could help AI clusters with becoming grid assets – absorbing surplus generation and reducing peak stress, even where average utilisation is modest. First modelling exercises by transmission system operators indicate that making large data centres flexible for roughly 5% of their operating time, or an equivalent share of annual demand, already unlocks many more grid connection points as well (P. Czyzak, 2026). Instead, current contracts and business models are likely to prioritise internal efficiency and service levels over system integration, leaving both idle capacity and flexibility unaddressed.

Siting and Permitting on Constrained Grids

The third bottleneck is spatial. EU level analyses show that in the long-term AI clusters are much more likely to concentrate in a handful of metropolitan and industrial regions, often on grids that are already constrained. AI clusters, especially those serving inference-heavy workloads, seek proximity to strong substations, fibre backbones and major user markets, and they benefit from local policy support and existing cloud ecosystems. The 2026 IDCA report estimates that more than half of European data centre capacity already operates in geographically concentrated clusters, frequently in areas where generation and network headroom are limited.

Against this backdrop, CADA’s proposal to create data centre acceleration zones is particularly significant. The idea is to concentrate support and streamlined permitting

in specific areas to speed up deployment across Member States. While the criteria proposed so far recognise the need to consider available and future power grid capacity in the designated area, they do not take into account energy bottlenecks stemming from hosting AI-optimised clusters.

In their current shape, CADA’s acceleration zones do not distinguish between AI-specific and traditional data-centre capacity, nor do they require assessments of the different electricity demand, load shape and flexibility profiles associated with particular workloads. This may encourage co-location of large AI clusters with less energy-intensive facilities without adequately accounting for their distinct effects on local grid capacity. Although the proposal can accelerate deployment through streamlined, zone-level environmental assessments and base-line permitting, and it encourages coordination with grid planning and the use of power purchase agreements, it still does not ensure that zones are selected or designed with attention to AI-specific demand profiles or credible flexibility arrangements. The result could be zones that are attractive from a digital industry perspective but risk being misaligned with energy system needs.

The AI Gigafactory tender offers a first concrete illustration of this risk. Its siting criteria are exhaustive on land control, zoning, and permitting status, geotechnical readiness, and the cluster’s own grid connection security (including a requirement to deliver 85–90% of contracted grid capacity to IT load). At the same time, the criteria include no scoring credit for choosing a site with documented available capacity in low-carbon generation or grid infrastructure, and the tender text does not reference CADA’s acceleration zones, the EED sustainability reporting framework, or the European Grids at any point. The tender’s phased rollout structure is nevertheless a step in the right direction, as it allows AIGFs’ capacity to be built in stages, but progression to full scale is not tied to evidence of local grid headroom, low-carbon power availability, utilisation or flexibility. Siting is therefore evaluated purely as a question of the AIGF’s own security supply, not of its fit within the wider electricity system.

Some of the Member States have already acted on the siting tensions prompted by constraints emerging in the context of traditional data centres. Ireland for example combines connection limits and sequencing rules for new data centres in constrained areas, while the Netherlands uses zoning to steer traditional data centres away from saturated nodes.

European-level measures, like the sustainability reporting framework for data centres established under the EED already aim to provide policymakers with basic energy performance and sustainability data stemming from traditional data centre operations. These data, although sparse due to the voluntary nature of the framework, now underpin the Commission’s rating scheme for

data centres, with the framework concentrating on generic efficiency and sustainability indicators, such as power and water usage effectiveness and renewable energy shares. The rating scheme does not introduce AI-specific metrics for load density, workload profiles or flexibility, nor does it tie such metrics explicitly to siting choices. Likewise, other initiatives such as the European Grids Package and the revised TEN-E Regulation do recognise large data centres as relevant grid users but stop short of defining dedicated rules for AI-optimised clusters. Without better alignment between CADA's acceleration zones, national grid development plans and local resource constraints, AI-specific capacity still risks being pulled into precisely the locations where it is hardest to integrate.

Policy Recommendations

The evidence in this brief points to **three core insights**:

- AI clusters are a distinct **type of industrial-scale load**;
- they are structurally **exposed to underutilisation**;
- and they currently **leave flexibility and siting questions largely unaddressed** from an energy system point of view.

Existing EU instruments already touch these issues, but they do not yet differentiate AI clusters clearly from traditional data centres. The recommendations below stay with-in the analysis presented and translate it into practical guidance for ongoing EU debates.

Recognise AI Clusters as a Separate Category

Building on the distinct characteristics of AI clusters that set them apart from traditional data centres – in hardware, workloads and load profiles – this brief has argued that treating AI infrastructure as a distinct category is a prerequisite for effective energy system integration across Europe.

The EU's AIGF tender exemplifies this through proposing first criteria specifically tailored to the prospective AI compute clusters. The result is mixed. The tender does treat the Gigafactories as a distinct facility type, with dedicated minimum thresholds for accelerator capacity (75,000–100,000 H100-equivalents) and IT load (120–150 MW) that go well beyond conventional data-centre categories.¹ However, it stops short of differentiating within that category: the same evaluation criteria apply regardless of whether a facility is training-heavy, inference-heavy or mixed, and no AI-specific load-profile or utilisation indicators are used to distinguish these workload mixes at the point of award.

In the implementation of the Data Centre Energy Efficiency Package, large clusters that host significant AI workloads should be identifiable as well, through specific indicators – a gap the AIGF tender itself does not fill, as it makes no reference to the data centre rating scheme.

National grid planning practice should explicitly include Member States and system operators treating AI clusters as industrial loads when assessing their impacts on generation, networks and adequacy.

Make Utilisation and Flexibility Explicit for AI Clusters

The analysis shows that AI workloads make AI clusters structurally prone to under-utilisation. At the same time, training and some batch workloads can, in principle, provide meaningful demand response. Utilisation and flexibility are therefore two sides of the same problem.

The Cloud and AI Development Act proposal already recognises improved utilisation and efficiency as a “Grand Challenge” for the Cloud and AI Leadership Initiative. The AI Gigafactory Call for Tenders published in July 2026 was the first opportunity to turn this Grand Challenge into binding criteria, but it did not fully do so: the tender includes no requirement for AIGF consortia to disclose expected utilisation trajectories or load factors. Grid-aware flexibility measures (time-shifting of training jobs, demand response participation, on-site storage) are all folded into a single sustainability sub-criterion, carrying only 0.4% of the overall technical evaluation score. For AI clusters that might benefit from EU-level support or accelerated treatment under CADA – including the future AI Gigafactories or comparable investments – this should translate into concrete expectations: projects should be required to disclose utilisation trajectories that reflect AI-specific workload patterns, and, where possible, they should identify which parts of their AI workloads can be exposed to demand response or other flexibility schemes.

The Commission's Data Centre Energy Efficiency Package, published in September 2026, introduces a proposed common EU rating scheme and label for data centres and launches preparatory work on minimum performance standards. While the package strengthens transparency on energy, water and selected grid-related indicators, it does not yet introduce AI-specific requirements for utilisation, workloads flexibility or location-sensitive grid compatibility. In implementing that package, policy-

¹ Under the EU's data centre reporting and sustainability rating scheme, a “very large” data centre is defined as one equipped with more than 10 MW of installed IT power.

makers should make use of AI-specific indicators – such as average IT load factor and basic load-profile measures – to monitor how intensively AI-relevant infrastructure is actually used and should link these metrics to participation in demand-response and flexibility mechanisms rather than treating utilisation as a purely internal efficiency concern.

Align Acceleration Zones and Siting Decisions with Energy Constraints

AI compute capacity in Europe risks becoming heavily clustered in a few metropolitan and industrial regions, often on constrained grids, and the proposed data centre acceleration zones under CADA currently pay limited attention to this reality. The AIGF tender confirms this pattern: its siting criteria extensively cover land control, permitting status and the facility's own grid connection security, but include no scoring credit for choosing locations with spare low-carbon generation or network capacity. The Commission's move to accommodate smaller AI compute clusters as part of the AIGF call is a step in the right direction, as it implicitly recognises integration limits, but does not yet translate into clear siting rules for AI workloads more broadly.

Therefore, criteria for designating data centre acceleration zones should explicitly embed the energy system considerations already discussed in this brief and favour locations with demonstrable headroom in low-carbon generation and network capacity. EU-level guidance on environmental assessment and permitting should reflect that large AI clusters behave as industrial actors and should be assessed accordingly, including their cumulative impact on electricity systems.

Conclusion

Policymakers' ability to connect the dots between tech sovereignty ambitions and energy realities is likely to become a litmus test of how fit the EU is to legislate in times of AI transformation. The AI Gigafactory tender criteria provided the first such test, but incorporated these conditions only partially. The political negotiations on CADA and the Data Centre Efficiency Package now represent the next, and arguably last near-term opportunity to embed conditions on utilisation, verifiable flexibility and siting into the rules that govern Europe's emerging AI clusters.

AI clusters can make or break Europe's energy transition. With the ongoing efforts to accelerate the uptake of data centre infrastructure across Member States, AI-specific facilities need to be recognised as distinct industrial actors on the grid. Supporting policies, stemming from this distinction, need to feature conditions on utilisation and verifiable flexibility, as well as appropriate guidelines on siting to effectively avoid a mismatch between technological ambitions and energy realities. At the same time, whether the projected expansion of AI infrastructure remains compatible with the EU's 2050 climate-neutrality objective will largely depend on how quickly new capacity is tied to additional low-carbon generation and grid reinforcement – an open question that current policy choices will decisively influence.

Bibliography

- Acun, F., Hankendi, C., Levine, E., Reynolds, H., Bardwick, J., & Coskun, A. K. (2026). Investigating Power Consumption Flexibility of AI Data Centers for Demand Response Participation. Available at: https://www.bu.edu/peaclab/files/2026/03/Flex-DC_Sim_ACM_E_Energy26.pdf
- A. Kaul (2026). The Great Datacenter Mismatch. Available at: <https://aisigint.com/blog/great-datacenter-mismatch>
- ASPI – Australian Strategic Policy Institute (2025). Abundant electricity isn't enough: China's overbuilt AI computing power is underused. Available at: <https://www.aspistrategist.org.au/abundant-electricity-isnt-enough-chinas-overbuilt-ai-computing-power-is-underused/>
- BloombergNEF (2026). *AI Data Centers, Energy and Finance: Dispatch from the BNEF Summit New York 2026*. Available at: <https://about.bnef.com/insights/clean-energy/ai-data-centers-energy-and-finance-dispatch-from-the-bnef-summit-new-york-2026/>
- Brockovich Data Center. Brockovich AI Data Center Reporting. Available at: <https://brockovichdatacenter.com>
- Chen, X., Wang, X., Colacelli, A., Lee, M., & Xie, L. (2025). Electricity Demand and Grid Impacts of AI Data Centers: Challenges and Prospects. Available at: <https://arxiv.org/html/2509.07218v3>
- Commins, J., & Irion, K. (2025). Towards planet proof computing: Law and policy of data centre sustainability in the European Union. Available at: <https://www.ivir.nl/publicaties/download/TechReg2025.001Commins.pdf>
- Czyzak, P. (2026). Data Center Flexibility (Intro). Available at: https://paczysak.substack.com/p/data-center-flexibility-intro?publication_id=7847607&post_id=194047574&isFreemail=true&r=7avnm
- Delahunt, J. (2026). Ireland's data centre policy reset: Decoding the latest policy developments shaping Ireland's data centre landscape. Available at: <https://kpmg.com/ie/en/insights/energy-utilities-telecoms/irelands-data-centre-policy-reset.html>
- EPRS – European Parliamentary Research Service (2025). Making Europe an AI continent. Available at: [https://www.europarl.europa.eu/RegData/etudes/BRIE/2025/775923/EPRS_BRI\(2025\)775923_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2025/775923/EPRS_BRI(2025)775923_EN.pdf)
- EpochAI (2026). Is AI compute crunch coming? Available at: <https://epoch.ai/gradient-updates/is-a-compute-crunch-coming>
- Eurostat (2026). Production of Electricity and Derived Heat by Type of Fuel. Available at: https://ec.europa.eu/eurostat/databrowser/view/nrg_bal_peh__custom_20322651/bookmark/table?lang=en&bookmarkId=7fd37d6c-0606-4c34-8b31-63fa5283e674&c=1772444282549
- European Commission (2025). AI Continent Action Plan COM(2025)165. Available at: <https://digital-strategy.ec.europa.eu/en/library/ai-continent-action-plan>
- European Commission (2025). A brief guide about the European AI Innovation Ecosystem. Available at: <https://euagenda.eu/publications/a-brief-guide-about-the-european-ai-innovation-ecosystem>
- European Commission (2026). Communication on European Tech Sovereignty and EU Open Source Strategy COM(2026) 503. Available at: <https://digital-strategy.ec.europa.eu/en/library/communication-european-tech-sovereignty-accompanying-eu-open-source-strategy>
- European Commission (2026). Proposal for a Regulation of the European Parliament and of the Council establishing a framework of measures for strengthening Europe's cloud and AI ecosystem (Cloud and AI Development Act), Article 10. Available at: <https://digital-strategy.ec.europa.eu/en/library/proposal-cloud-and-ai-development-act-cada>
- European Commission (2026). Annexes to the Regulation of the European Parliament and of the Council establishing a framework of measures for strengthening Europe's cloud and AI ecosystem (Cloud and AI Development Act), Article 1, para 2. Available at: <https://digital-strategy.ec.europa.eu/en/library/proposal-ai-cloud-and-ai-development-act-cada>
- European High Performance Computing Joint Undertaking (EuroHPC JU), Call for Tenders for the Selection of Artificial Intelligence Gigafactory Consortia and the Establishment of AI Gigafactories: Tender Specifications, Ares(2026)7435477. Available at: https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/tender-details/docs/bc712395-692d-41ab-aa25-8e254ae86eb2-CN/CFT%20AIGF%20-%20Tender%20Specifications_V1.pdf
- Hinterholzer, S., et al. (2025). Assessment of Next Steps to Promote the Energy Performance and Sustainability of Data Centres in EU, Including the Establishment of an EU-Wide Rating Scheme—First Technical Report. Available at: <https://euagenda.eu/publications/assessment-of-the-energy-performance-and-sustainability-of-data-centres-in-eu>
- IEA – International Energy Agency (2026). Key Questions on Energy and AI. Available at: <https://iea.blob.core.windows.net/assets/3179f7f8-01f6-4dd6-bffa-c9f7b73f1dc9/KeyQuestionsonEnergyandAI.pdf>
- IDCA – International Data Center Authority (2026). Global Data Center Report 2026. Available at: [https://firebasestorage.googleapis.com/v0/b/idca-f9873.appspot.com/o/Research%2FGlobal%20Data%20Center%20Report%20\(2026\).pdf?alt=media&token=66b00a05-d523-4a43-9363-b4ad49249d02](https://firebasestorage.googleapis.com/v0/b/idca-f9873.appspot.com/o/Research%2FGlobal%20Data%20Center%20Report%20(2026).pdf?alt=media&token=66b00a05-d523-4a43-9363-b4ad49249d02)
- Mann, T. (2025). Google Agrees to Pause AI Workloads to Protect the Grid When Power Demand Spikes. Available at: https://www.theregister.com/2025/08/04/google_ai_datacenter_grid/
- McArdle, J., & Terras, P. (2025). System overload: How new data centres could throw Europe's energy transition off course. Available at: https://beyondfossilfuels.org/wp-content/uploads/2025/02/SystemOverload_Report.pdf
- North American Electric Reliability Corporation (2026). NERC Issues Level 3 Alert, Reliability Guideline Focused on Large Load Challenges. Available at: <https://www.nerc.com/newsroom/nerc-issues-level-3-alert-reliability-guideline-focused-on-large-load-challenges>
- Nowicka, M. (2026). From Chips to Grids. Why Energy Constraints Influence AI Compute Expansion in Europe. Available at: <https://www.interface-eu.org/publications/ai-compute-energy-bottlenecks#>
- SemiAnalysis (2026). Inside the 800VDC Revolution – Part 1. Available at: <https://newsletter.semianalysis.com/p/inside-the-800vdc-revolution-part>
- SemiAnalysis (2026). US Grid Constrains: Towards 40GW+ of Behind-The-Meter Datacenter by 2028? Available at: <https://newsletter.semianalysis.com/p/us-grid-constraints-towards-40gw>
- Sheng, Y., Zhang, C., Zhu, Z., Xu, H., Wen, J., Wang, R., Yang, J., Wang, Q., & Bu, S. (2026). Power for AI Data Centers: Energy Demand, Grid Impacts, Challenges and Perspectives. *Energies*, 19(3), 722. Available at: <https://doi.org/10.3390/en19030722>
- State of Maine, House Democrats (2026). Sachs issues statement on governor's decision to veto bill establishing temporary pause on data centres. Available at: <https://www.maine.gov/housedems/news/sachs-issues-statement-governors-decision-veto-bill-establishing-temporary-pause-data-centers>
- Commission Delegated Regulation (EU) 2024/1364 of 14 March 2024 on the first phase of the establishment of a common Union rating scheme for data centres. Available at: https://eur-lex.europa.eu/eli/reg_del/2024/1364/oj/eng
- Virkkunen, H. & Jørgensen, D. – European Commission (2026). EC press conference by European Commission Executive Vice-President Henna Virkkunen and Commissioner Dan Jørgensen on the Tech Sovereignty package, 3 June 2026. Available at: <https://audiovisual.ec.europa.eu/en/ebs/grid?ebs=yes&ebsplus=yes&date=20260603>

About the author

Maria Nowicka is a Policy Researcher at the Global Chip Dynamics Programme at [interface](#). Her work examines how semiconductor supply chains and AI compute infrastructure shape Europe's economic resilience, technological sovereignty and green transition. mnowicka@interface-eu.org

interface

interface is a European think tank specialising in tech public policy. Interface's work aims to ensure that political decision-makers have the expertise necessary to create policies that put the public interest first.

Powering AI: How Europe can align its AI data centre buildout with energy realities

AI infrastructure build-out is increasingly shaping the EU's tech-sovereignty agenda and the electricity systems that underpin it. Unlike conventional data centres, AI clusters concentrate dense accelerator hardware, latency-sensitive workloads, and often need firm capacity near demand hubs. This brief argues that Europe's AI expansion will require dedicated energy planning, and that such planning is still largely absent in proposed legislation. Without it, the continent risks locking in vulnerabilities that could undermine both its digital and climate neutrality ambitions.

Further information on the topic can be found here:

➤ justclimate.fes.de