

Leevi Saari and Boxi Wu
May 2026

Chasing the AI Cloud in Europe

*Handover Blindness and Implications
for EU AI Sovereignty*



Imprint

Published by

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

Issuing Department

Competence Centre on the Future of Work
Cours Saint Michel 30a, 1040 Brussels, Belgium

Responsibility for Content and Editing

Justin Nogarede
justin.nogarede@fes.de

Proofreading

Lloyd Bingham, Capital Content

Design/Layout

pertext | corporate publishing
www.pertext.de

Cover picture

lolipop/stock.adobe.com

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.

May 2026

© Friedrich-Ebert-Stiftung e.V.

ISBN 978-3-98628-877-8

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

➤ www.fes.de/publikationen

Leevi Saari and Boxi Wu
May 2026

Chasing the AI Cloud in Europe

*Handover Blindness and Implications
for EU AI Sovereignty*

Contents

Executive Summary	3
1. Introduction	5
1.1 From traditional to AI cloud	6
1.2 Geographies of computation: where are the data centres?	9
1.3 The emergence of the alternative AI cloud market: the search for computational alternatives	12
2. Alternative AI cloud providers and their strategies in the EU market	15
2.1 Market actors	15
2.2 Market strategies	16
3. Challenges for the EU AI Cloud	19
3.1 Physical constraints and bottlenecks	19
3.2 Demand concentration and the market risk	20
3.3 Supplier concentration and the Nvidia risk	20
3.4 Market structure challenges	22
3.5 Industrial policy for hyperscalers? The case of EIF, the CDP and Apto's Lacchiarella site	23
4. Policy implications: questions for EU AI policy	26
5. Conclusions	28
References	29

Executive Summary

The European Commission plans to transform Europe into an ‘AI Continent’, both to increase the EU’s technological sovereignty as well as its competitiveness. One key part of this objective is the role of material infrastructure. Consequently, the bolstering of European artificial intelligence (AI) compute capacity is a matter of urgent geopolitical interest, with the European Commission’s AI Continent Action Plan aiming at tripling data centre capacity in the EU by 2033.

The policy ambition for technological sovereignty is clear, yet the roadmap for implementation remains precarious. Policy discussions often rely on the metaphor of ‘the cloud’, which masks critical technical distinctions. **Conventional cloud infrastructure**, which powers standard IT, differs fundamentally from the **AI cloud** required for large-scale model training and deployment. Currently, EU initiatives focus almost exclusively on public-private High-Performance Computing (HPC) clusters for model **training**. This focus overlooks the most critical phase of the AI lifecycle: **inference** (the ongoing usage of models).

This creates a phenomenon we call **‘handover blindness’**: large models are trained on European public infrastructure (e.g. EuroHPC Leonardo or LUMI) but are then ‘handed over’ to American hyperscalers (Amazon, Google and Microsoft) to provide the necessary compute for inference. Value capture remains entrenched within ecosystems dominated by the US hyperscalers. As model capabilities saturate, the relative importance of training will decrease while inference demand grows, potentially rendering public investments in large-scale generative AI training-optimised computing infrastructure partially misallocated.

To break this dependency, a new ecosystem of European AI compute providers is emerging as a potential alternative to provide AI compute for European companies. This ecosystem is the focus of the paper. Specifically, it analyses neoclouds: private, AI-native compute providers. **The analysis includes a sample of 19 European AI cloud providers and identifies four primary market entry strategies that they pursue:**

- **Consolidation:** Aggregating smaller compute providers to achieve scale.
- **Specialization:** Serving highly regulated sectors (defence and healthcare) with sovereign requirements.

- **Upgrading:** Moving beyond raw compute to offer high-value services.
- **Marketplace:** Providing flexible, on-demand capacity via specialised platforms.

As the analysis brings to the fore, the expansion of European AI cloud capacity, in the service of increased technological autonomy and competitiveness, **faces several structural challenges that require explicit policy attention:**

- **Physical constraints** regarding land and water use, power generation, and grid transmission. **The EU’s objective** of “tripling data centre capacity” is primarily limited by these physical constraints rather than just capital or hardware availability. They are driving a geographical reorganization of European infrastructure, with compute centres moving toward locations with cheaper land and energy, such as the Nordics or Iberia.
- While US hyperscalers largely finance their capital expenditures out of massive existing cash flows, most European neoclouds rely on debt secured by collateral structures anchored in their GPU assets. These **financial constraints** make them vulnerable to demand fluctuations, interest rate hikes, and any cooling of the “AI bubble”. This financial fragility suggests that a significant consolidation phase is likely, which may leave only a handful of viable European providers standing.
- **The high amortisation costs of AI hardware require near-constant utilisation to remain profitable.** The demand from public or regulated industries, such as defence and healthcare, that form the primary target market for many alternative providers, is insufficient to amortize these costs. As a result, many of the largest neoclouds function as “overflow capacity” for the hyperscalers they intend to compete with. This dependency creates a precarious business model where the survival of European alternatives is dictated by the spillover demand of their primary competitors.
- **Nvidia is a pivotal partner and liquidity provider for Europe’s neocloud ecosystem** and has indirectly become a geopolitically salient partner for the EU’s technological sovereignty aspirations. Yet its deep financial ties, ecosystem programmes, competing DGX Cloud marketplace and rapid hardware cycles create **structural dependencies and economic tensions** for neocloud pro-

viders. Nvidia's product roadmap prioritizes rapid hardware cycles. This directly undermines the financial viability of neoclouds, which require longer periods to pay down the debt used to purchase that very hardware.

- **Neoclouds are also facing a 'commodity trap' in their attempt to move toward higher value-added services**, currently squeezed between Nvidia, which controls the hardware and much of the low-level software stack (CUDA), and the hyperscalers, who control the primary distribution networks and downstream value-added services.
- **Neoclouds compete within an ecosystem where the incentives of other stakeholders are stacked against them.** Across the AI value chain, actors gravitate toward incumbents: end-users are locked in by free credits and skill-path dependencies, while model providers prefer hyperscalers for their built-in distribution networks. Similarly, data centre and private credit providers favour the stability of hyperscalers over the riskier, asset-backed lending models of neoclouds. This challenge was recently exemplified by the European Investment Bank's financing of a hyperscaler project in Italy, highlighting the disconnect between high-level digital sovereignty goals and ground-level financial execution.

This analysis highlights that the AI compute value chain is structurally tilted toward incumbent cloud infrastructure providers and value chains. Even existing alternative providers are deeply entangled with Nvidia, which is entrenching its role in both public and private EU compute initiatives and locking in existing technical and economic path dependencies. This raises significant concentration risks and competition concerns, possibly tethering European AI policy to monopolies and hyperscaler paradigms of large-scale AI.

In the worst-case scenario, the pursuit of technological autonomy ends up subsidising AI cloud providers through policies that streamline permitting, land use and energy policies. This is likely to benefit foreign incumbents with painful trade-offs for European policy preferences on autonomy, the green transition, land use policies and the control over AI technological trajectories. Supporting policy autonomy and preventing goal displacement requires strategic interventions that account for the dynamics of the AI compute ecosystem.

Policy considerations

- **To sustainably expand European AI infrastructure**, policy should prioritise energy and grid investments before the acquisition of hardware. As the most durable elements of AI infrastructure, they can be built on a 'no-regrets' foundation; even if AI demand fluctuates, these assets are easier to repurpose to benefit the wider European economy. Integrating energy considerations into the initial policy framework also avoids a scenario where the consequences of data centre buildout (increased costs, distributional concerns, and environmental impacts) are presented as fait accompli without upstream political discussion. Moreover, before expanding data centre capacity, it is important to consider to what extent the current capacity is used efficiently.
- **The public interest is not synonymous with an 'Nvidia arms race'.** Recent advancements in compute-constrained AI models like OpenAI's o3-mini or DeepSeek's R1 models point towards the feasibility of more cost-effective and sustainable modes of AI development. **Policymakers should be mindful of the implications of the obsolescence cycle of the current chip architectures for public procurement. Procuring behind the depreciation curve, i.e. buying chips slightly behind the bleeding edge would lead to substantial cost-savings and maintain political autonomy.** Furthermore, a strategic diversification of chip providers is essential to grant the public sector agency needed to prevent vendor capture and maintain long-term technological autonomy.
- **To prevent market fragmentation and capture by the existing incumbents, compute demand and supply must be pooled on a centralised platform.** Currently, European providers are fragmented, with many competitors competing over little demand. This makes them vulnerable to becoming drawn into the gravitational pull of hyperscaler demand, or becoming commodified providers as part of private aggregators such as Nvidia DGX Lepton. As a consequence, it is important to explore demand-side coordination, competitive unbundling of cloud markets and controlled integration through interoperability to counter the trend towards concentration.

1. Introduction

In response to rising geopolitical tensions, increasing European sovereignty has moved to the forefront of the policy agenda. Under European Commission President Ursula von der Leyen, digital sovereignty became the centrepiece of the Commission's digital agenda. As part of this push, efforts focused on cloud infrastructure, as the 'computational underbelly' of the modern economy, were driven by concerns over remaining competitive and offering enhanced data protection. However, the heightened geopolitical risks associated with the second Trump administration have shifted the European policy window: from curtailing foreign actors through regulation towards building domestic capacity and alternative infrastructures.

The emergence of the large-scale artificial intelligence (AI) paradigm following the release of ChatGPT in 2022 spurred corporate and political excitement. Combined with intensifying US-China technological competition, governments around the globe have reframed digital sovereignty concerns through the lens of AI technologies and their underlying infrastructure (Hawkins, Lehdonvirta & Wu 2025). Consequently, the cloud infrastructure underpinning the AI revolution, – known as the AI cloud and defined by the OECD as “the specialised hardware and software stacks required to train and run AI models” – has emerged as a key springboard for digital sovereignty discourses and policy initiatives in recent years.

This has led to new forms of AI industrial policy (Kak & Myers West 2024), including the procurement of chips required for AI development and deployment,¹ which are housed in data centres: physical warehouses that host the infrastructure required to operate and power AI chips (OECD 2025a). In the past few years, the European Commission has launched several initiatives to build these capacities to protect European AI sovereignty. These include: support for AI Factories that upgrade the existing European High-Performance Computing (EuroHPC) capacity; the building of several Gigafactories already announced, the EU AI Continent Action Plan and the upcoming EU Cloud and AI Development Act, with an objective being to “triple the data centre capacity in Europe” (European Commission 2026). This points to the evolving landscape of AI infrastructure in the European Union.

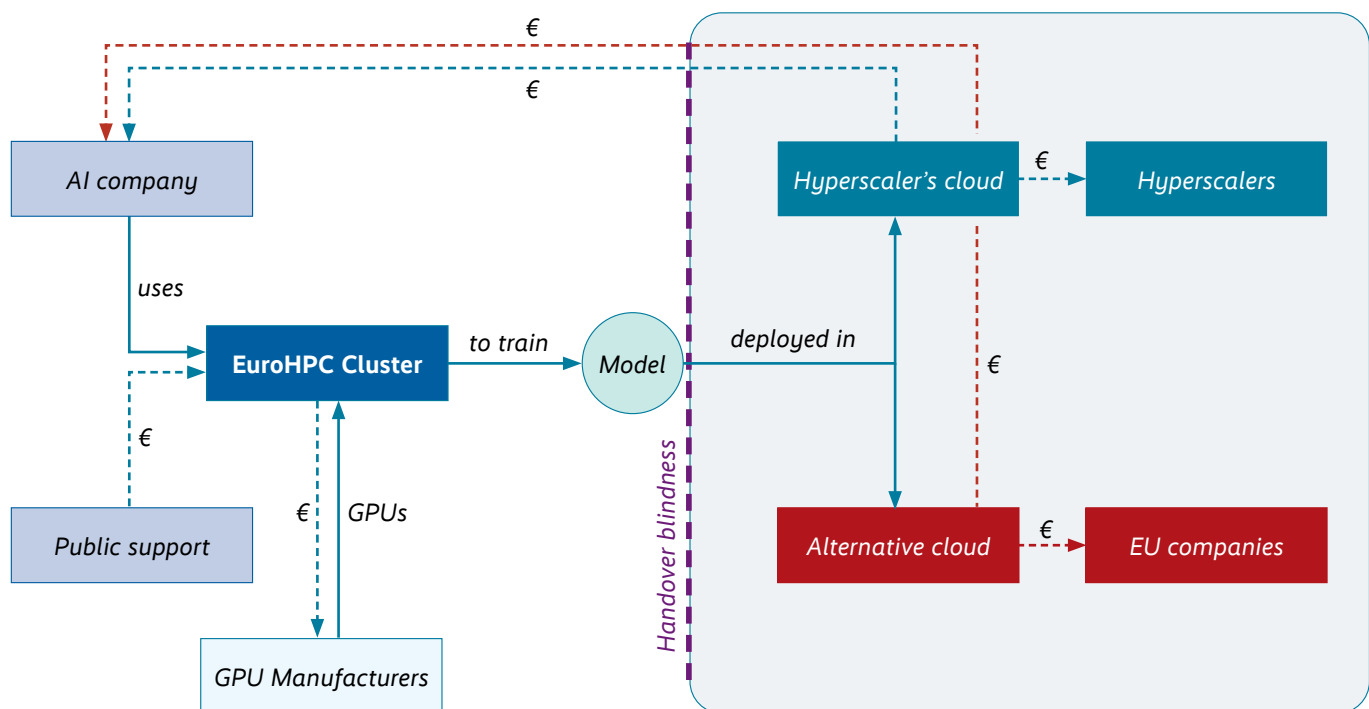
Historically, European AI compute policy has focused almost exclusively on the training phase. While the EuroHPC and AI Factory initiatives primarily address the computational bottlenecks of large-scale **training** runs used to create the models, they neglect the ecosystem where these models are actually deployed (**inference**), leading to what we term **handover blindness** in European AI compute policy. This is a notable omission. First, much of the value generation and extraction of AI models happens in the inference phase. Second, industry trends point almost exclusively to the increasing importance of inference services as the driver of future AI compute demand. Currently, this inference capacity is dominated by American hyperscalers. For European policymakers, this creates a potential paradox: a large AI model trained on a European public supercomputer will likely remain reliant on a US-controlled cloud for its daily operations, possibly negating the sovereignty gains of the training phase.

Consequently, there is a substantial policy interest in finding alternatives to US hyperscalers for inference. We have identified three groups that are positioning themselves to capture the revenue resulting from the push towards sovereign architectures due to shifting geopolitical tensions. One notable emerging group has been **neoclouds**: specialised AI compute providers emerging in response to rising computational demands. There are also pivoting **conventional cloud operators** who are expanding their portfolios to include AI compute offerings. Lastly, European **telecommunications operators** are looking to provide computational capacity. In addition to hyperscalers adjusting their operations to include novel sovereignty guarantees, these form the set of potential providers of AI inference compute in the European Union.

In this report, we focus primarily on **neoclouds**. While neoclouds could be expected to serve as potential bastions of European digital sovereignty, their political economy remains poorly understood. There is a lack of data on their market strategies, supply-chain dependencies and actual capacity to compete. This policy brief addresses these gaps by analysing the emerging interaction of political and economic forces in the EU's AI cloud market and its implications for long-term EU technological autonomy.

¹ By “AI chips” we refer to specialised parallel processing chips that are used to train and run large deep learning models. Examples include graphic processing units (GPUs, e.g. Nvidia and AMD), tensor processing units (TPUs, e.g. Google Ironwood), language processing units (LPUs, e.g. Groq) and wafer-scale engines (WSE, e.g. Cerebras). Currently, the vast majority of AI chips are GPU-variants, with recent signals on the emergence of TPUs as an alternative framework after notable models (Gemini 3) were successfully trained on those, and they are served to external parties as well.

The handover problem in the European AI compute ecosystem



Source: Authors, based on Arun 2025

1.1 From traditional to AI cloud

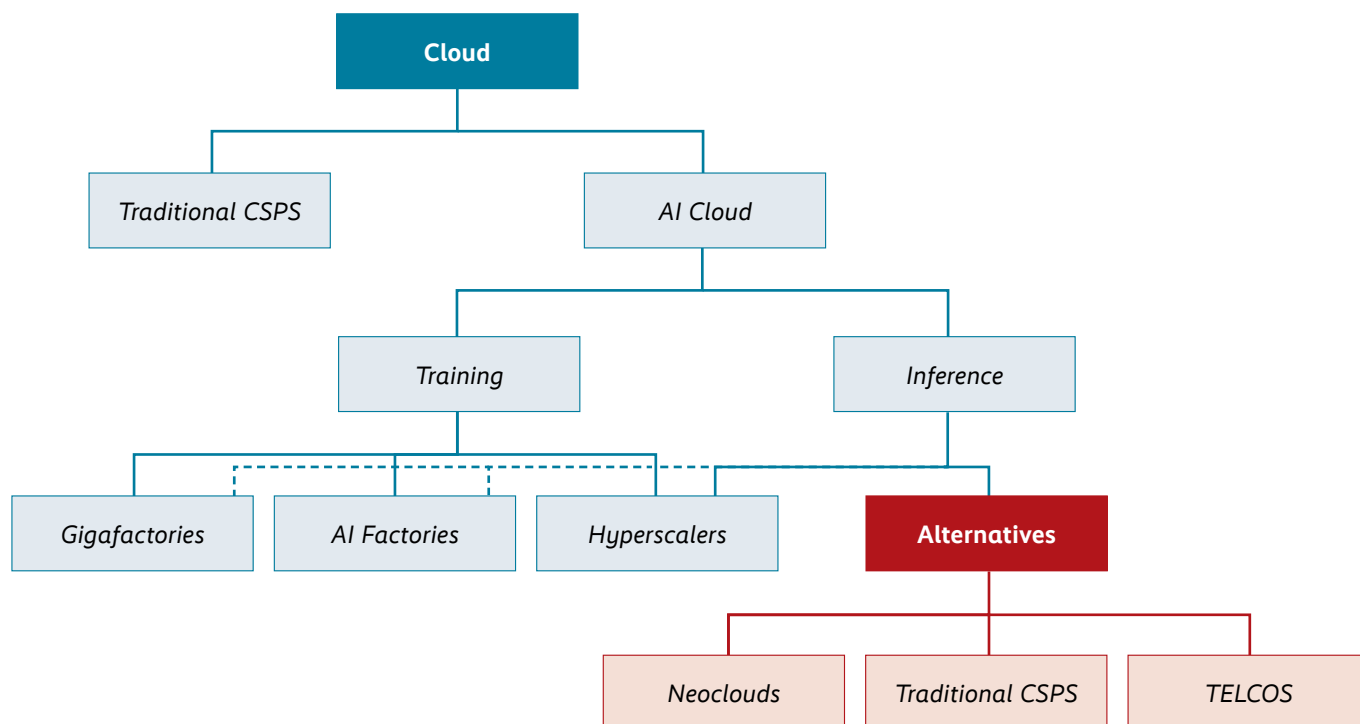
To start with the basics: **what is “the cloud”**? Contrary to the imagined seamlessness of the digital cloud (Hu 2016), cloud computing infrastructure forms the definite physical backbone of the contemporary digital economy. It comprises the software and hardware required to store and process data, including the development and deployment of AI models and downstream applications. This infrastructure includes computer chips, which are accessed remotely via the internet but stored in data centres (physical, warehouse-like facilities that house the software, hardware and networking equipment required for data storage and processing) (OECD 2025a). Computational resources are hosted in data centres and either accessed directly by companies (e.g. Google directly accesses computations via a data centre) or as a shared on-demand service that individual customers access via a cloud platform (e.g. a developer accesses computation via the Google Cloud Platform) (Lehdonvirta, Wu & Hawkins 2024). Cloud can be further broken down into access to infrastructure to run arbitrary software (Infrastructure-as-a-Service), access to an environment to develop applications via the cloud (Platform-as-a-Service) or ready-to-use software accessible via the cloud (Software-as-a-Service) (OECD 2025b).

When discussing digital sovereignty, it is vital to distinguish between the **conventional cloud**, referring to the broad market supporting general data processing (health-

care, finance, etc.), and the specialised **AI cloud** used for running AI workloads. Both conventional cloud companies and AI cloud companies own the computational resources (central processing units, or CPUs) required to store and process data, held in specialist infrastructure. Cloud companies in Europe, of which 70% are US hyperscalers and 15% are European companies such as OVHcloud, Hetzner and IONOS Cloud (Verbovnikov 2026), support many industries, such as healthcare, financing and manufacturing, by enabling data storage and processing.

However, the AI cloud requires data centres with higher-capacity AI chips that run specific computational operations for AI. These chips for AI workloads are extremely expensive and require a higher rack density (amount of energy consumed by a single server rack) compared with traditional CPU and storage servers required for general purpose computing: traditional data centres require rack densities of 12–15 kW per rack compared with 100–150 kW per rack for AI data centres, with newest models expected to need up to 600 or 900 kW (Patel, Nishball & Eliahou Ontiveros 2024; Walton 2025). AI chips also require specialised supporting infrastructures with cooling technology (e.g. direct-to-chip or immersion cooling) to keep racks cool, compared with traditional air cooling methods (Computer Room Air Conditioning, or CRAC) used in traditional data centres. We have summarised the key differences between the conventional and AI cloud in Table 1.

The typology of the cloud(s)*



* The extent to which AI Gigafactories and AI factories can be used for inference is an ongoing policy discussion. Source: Authors, based on Arun 2025

Table 1

Conventional versus AI Cloud

	Conventional Cloud	AI Cloud in the European Union
Companies	<ul style="list-style-type: none"> → Hyperscale cloud providers e.g. Amazon, Microsoft, Google → Conventional cloud providers 	<ul style="list-style-type: none"> → Hyperscale cloud providers → (Pivoting) conventional cloud providers → Telecom operators → Neocloud providers e.g. CoreWeave
Workloads	<ul style="list-style-type: none"> → Data storage and processing e.g. web-hosting, cloud storage, data analysis → Multi-use 	<ul style="list-style-type: none"> → AI training and inferencing, including data storage and processing → Tailored to AI intensive workloads
Hardware	<ul style="list-style-type: none"> → Traditional chips (CPUs) → Ethernet cabling 	<ul style="list-style-type: none"> → High-end accelerator chips, e.g. Nvidia GPUs, Google TPUs and chips with high-bandwidth memory. → Advanced networking
Cooling	<ul style="list-style-type: none"> → Air-based cooling (CRAC) 	<ul style="list-style-type: none"> → Liquid cooling (direct-on-chip or immersion)
Data centres	<ul style="list-style-type: none"> → Rack density typically 5–10 kW power per rack → Overall power capacity typically less than 100 MW 	<ul style="list-style-type: none"> → Rack density typically above 5–10 kW power per rack, with some Nvidia B200 120 kW, 125–150 kW for AMD MI355x and even 600 kW for Rubin series → Overall power capacity of a data centre typically greater than 100 MW
Key location factors	<ul style="list-style-type: none"> → Proximity to Internet Exchange Point (IXP), land, network effects 	<ul style="list-style-type: none"> → Availability of cheap land and power
Market depreciation of the hardware	<ul style="list-style-type: none"> → Linear (5–10) 	<ul style="list-style-type: none"> → Accelerated (2–6)

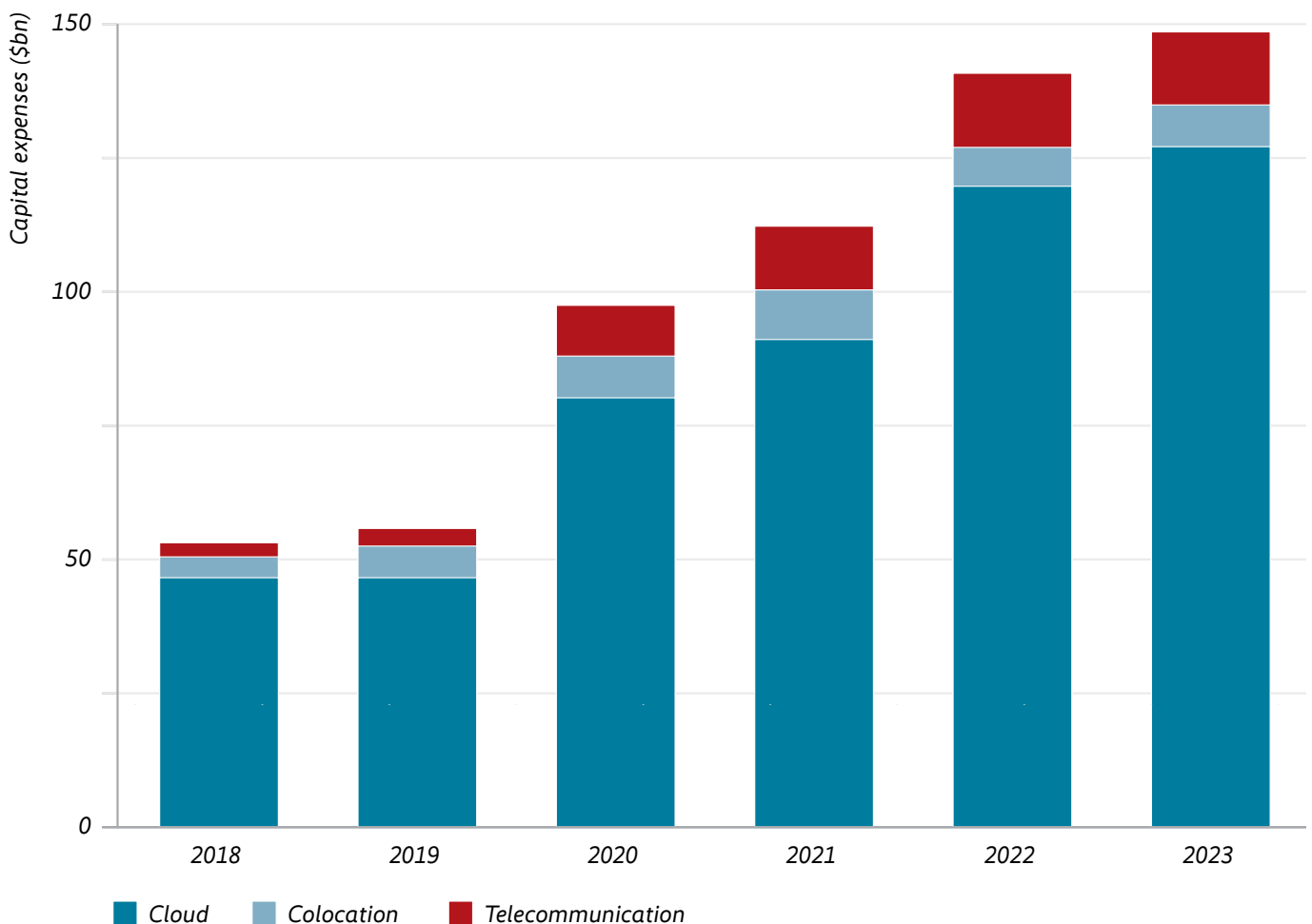
Key providers in global compute provision include US hyperscalers and companies that specialise in data centre development and operation. Some companies, such as Apple or Meta, build computational capacity exclusively for their own use, whereas others rent computational capacity to other companies (e.g. Amazon Web Services, Google Cloud and Microsoft Azure). In addition, **colocation providers** build and lease data centres for end users (e.g. Digital Realty, Equinix, CyrusOne and NTT). While their customers are confidential, these colocation firms use three main business models: hyperscale (leasing to one actor), colocation wholesale (leasing to several companies) or colocation retail (leasing on demand). Finally, **telecommunications operators, such as Deutsche Telekom and Telefonica**, increasingly operate their own data centres. These three actors (cloud, colocation and telco) have traditionally composed the majority of data centre capital expenses, as can be seen in the Figure 3 below.

Both traditional and AI cloud markets are concentrated. Over the past decade, US and Chinese hyperscale

cloud providers have emerged as dominant. Conventional cloud, used to store and process global internet activities, has a vast physical footprint of data centres (Lehdonvirta, Wu & Hawkins 2023). According to the Uptime Institute, Amazon Web Services (AWS), Microsoft Azure and Google Cloud controlled around two-thirds of global cloud spending in 2023, a notable increase from 47% in 2016 (Rogers 2025). There are also smaller and specialised cloud providers, such as OVH Cloud, Oracle and IBM Cloud. Systematic data on the size, geography and actors of the AI specific cloud market is currently unavailable, requiring a case-by-case analysis. While AI cloud provision is dominated by hyperscaler cloud providers, particularly US and Chinese companies, there is an emerging market for neocloud providers that specialise in AI compute provision, as well as smaller conventional and telecoms cloud providers that are beginning to offer specialised AI compute (further definitions in Section 2 below). This longer tail of AI cloud providers is the focus of this report.

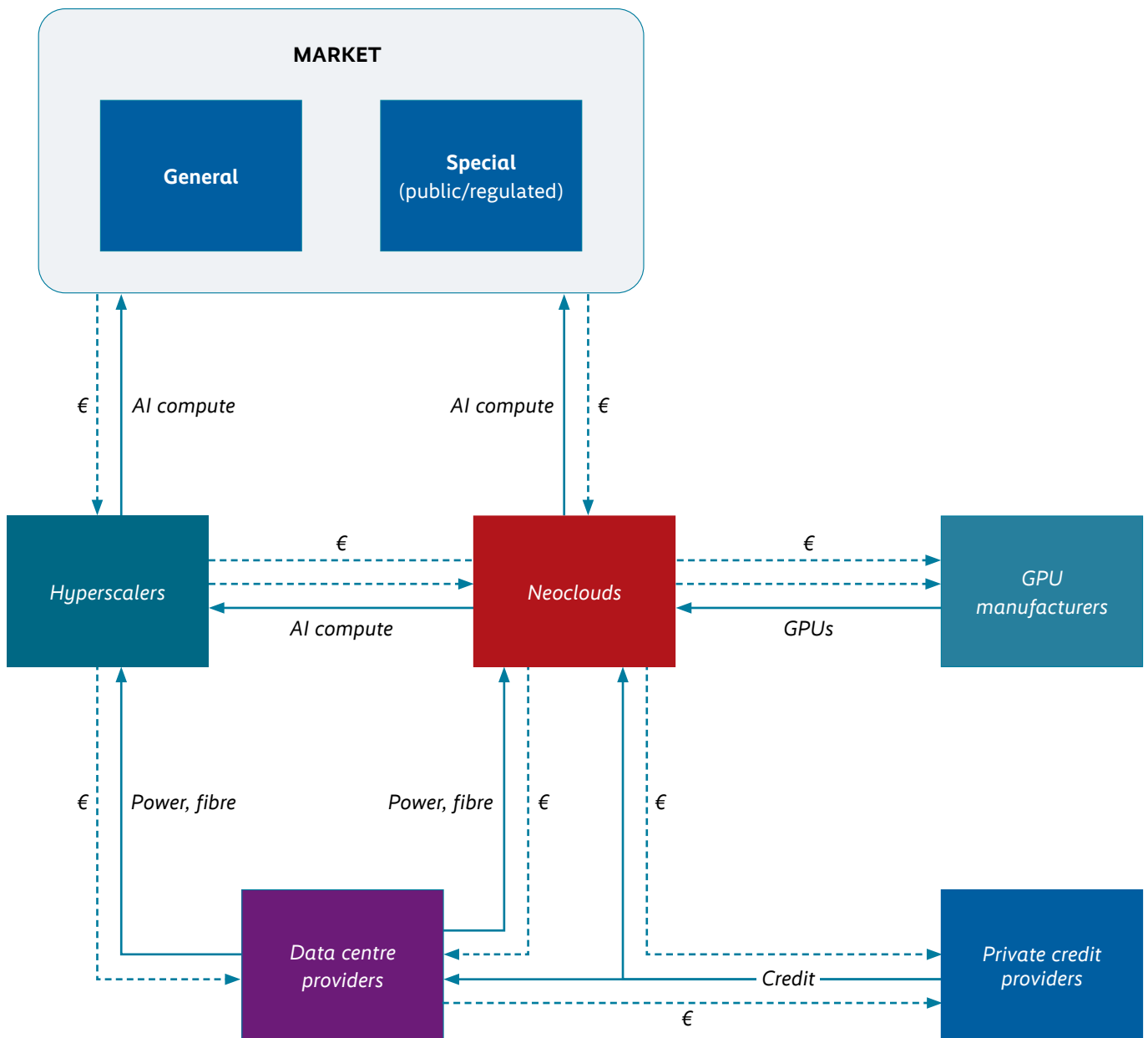
Data centre capital investments per type (2018–2023)

Figure 3



Source: Data from Omdia (a technology research and advisory group; see: <https://omdia.tech.informa.com/>), 2024.

The structure of the AI cloud market



Source: Authors, based on Arun 2025. Note that sometimes, hyperscalers rely on financial intermediaries for the financing of their data centers. In addition, a key point is that the market does not exist as an undifferentiated ‘arms-length’ market but rather is segmented into already-existing distribution networks. These create structural advantages for hyperscalers through economies of scale and integration benefits.

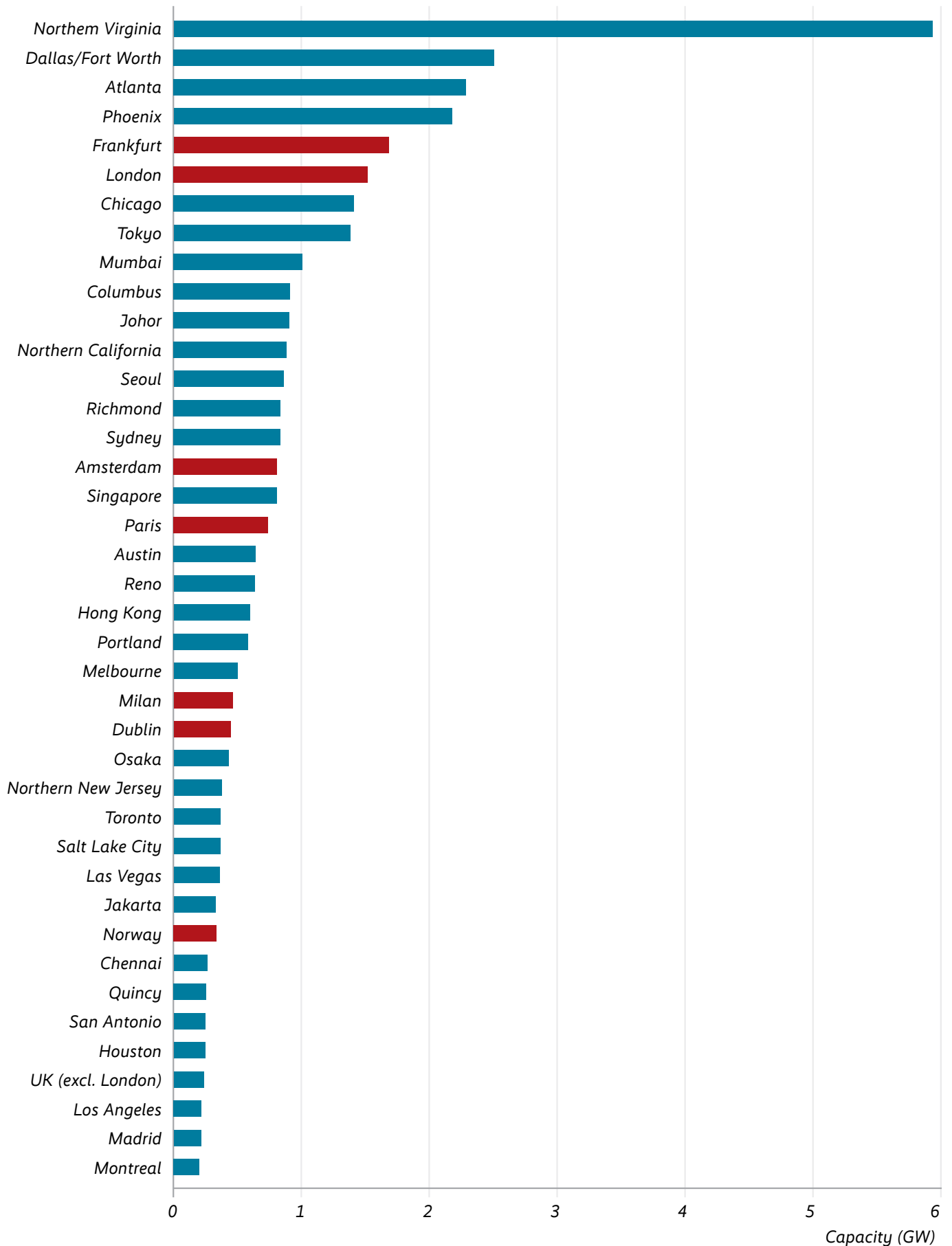
1.2 Geographies of computation: where are the data centres?

The global geography of traditional data centres is determined mainly by physical factors, such as the availability of power and the presence of high-speed data interconnections.² The global hotspot for data centres is Northern

Virginia, where the historical MAE-East interconnection created a path dependency thanks to the availability of interconnects (Zografos 2025). Moreover, data centre areas benefit from network effects, where the presence of other connections guarantees the lowest possible latency, which entrenches existing path dependencies. This has historically created concentrated data centre hubs.

² The International Energy Agency hosts a useful dashboard to see connections between energy and AI. see <https://www.iea.org/data-and-statistics/data-tools/energy-and-ai-observatory?tab=Energy+for+AI>.

Global data centre power capacity by region (Q3/2025)



Source: Datacenterhawk

In Europe, the key data centre regions have been Frankfurt, London, Amsterdam, Paris and Dublin, known as the FLAP-D market. Frankfurt hosts the DE-CIX internet exchange point, London hosts a plethora of high-speed financial trading houses and Amsterdam has another exchange point (AMS-IX). Paris also sits as the heart of key interconnections and Dublin has benefitted from the presence of the American hyperscalers (Boonstra 2023). According to the real estate investor CBRE, London and Frankfurt account for approximately half of total data centre supply in FLAP-D at the end of 2025, estimated to be 1.3 GW and 1.2 GW markets, respectively (CBRE 2025a).

The market for European cloud capacity is highly saturated. High demand, as well as supply constraints, are demonstrated by the vacancy rate remaining below 10% in 2025 (CBRE 2025a). The buildout of new data centres has currently been severely curtailed in the key locations due to considerations over energy and land use. According to some commentators, 33 to 45% of electricity in Frankfurt is spent on data centres, while in Dublin, estimates range from 64% to 95% (Gröger et al. 2025). This has halted data centre construction in some markets,

leading to spillovers to other places (Madrid, Warsaw, Milan and Zurich) (CBRE 2025b) with more abundant energy, planning permits, land availability or ambient cooling to help control the costs of running the data centres (CBRE 2025c).

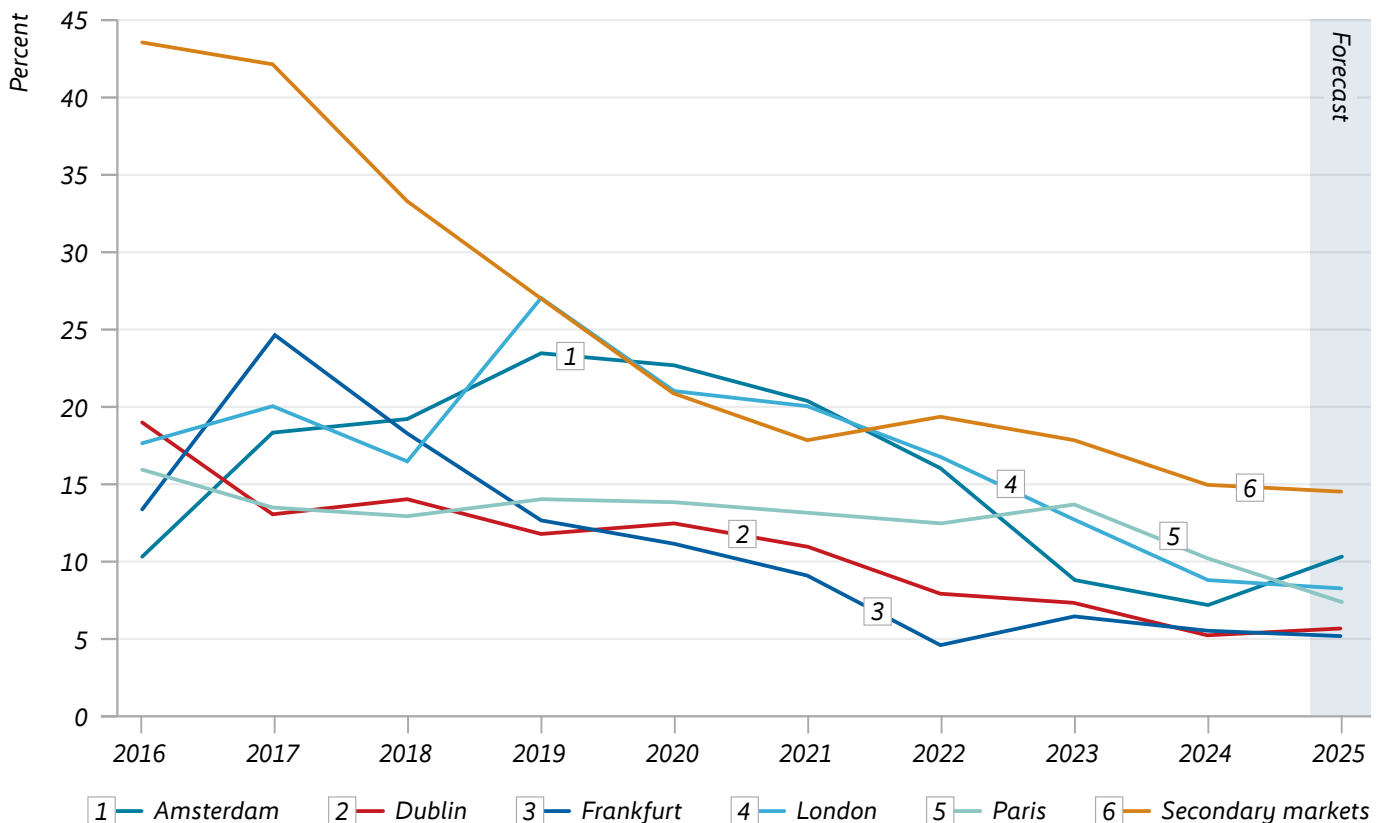
AI infrastructure requires a wholly different size of data centre capex in terms of geography, energy and deployed capital (CBRE 2023). Geographically, the large AI data centres are built outside metropolitan areas and conventional cloud hotspots because of cheaper greenfield land, access to electricity and reduced local congestion. This has implications for energy demand: while a 10 MW data centre was previously considered large and a 100 MW centre a hyper-scale project, references to projected 1,000 MW (1 GW) sized projects are becoming commonplace.

For example, Project Rainier by AWS, currently being built in Indiana to serve AWS AI workloads, is planned to be 2.2 GW when operational. The xAI Colossus data centre draws 150 MW of power and houses 100,000 AI chips³, in the range of AI chip clusters forecasted to be built under the EU's Gigafactories initiative. At their limit, aspirational

³ The Colossus 2 is planned to draw 2 GW of power when ready.

European cloud computing market vacancy rates in MW, 2016–2025 (Q3, 2025)

Figure 6



Source: CBRE, 2025b.

projects such as Stargate, the computation infrastructure project between OpenAI and Microsoft, as well as the Meta Hyperion 5 GW project in Louisiana, would single-handedly draw a maximum 5 GW of power.⁴

These infrastructure projects are orders of magnitude larger than previous iterations of the cloud. This scale difference poses challenges for the energy grids and power capacity as they stand, with some of the data centres building unpermitted natural gas and oil domestic power plants ‘behind the wire’ to power these needs. The shortage of energy components has even led to some actors using repurposed aircraft engines to power the new data centres (James 2025).

The investment required is also an order of magnitude higher than for conventional cloud providers.⁵ According to some estimates, 1 GW of AI compute required \$50 bn of capital expenses investment, depending on various factors such as land availability and energy costs, which has led to a record \$400 bn capital expenses on AI infrastructure in 2025 (Bobrowsky 2025). While much of the AI buildout has been funded by the dominant tech platforms, even hyperscalers have been forced to resort to external financing to manage the balance sheet implications of this scale of buildout.

This new AI cloud has driven a shift in the traditional geography of data centres. In the United States, rural locations in states such as Texas, Indiana and Louisiana have become popular owing to the availability of cheap, level land and a relaxed approach to environmental permitting. In Europe, Nordic countries including Norway and Iceland have emerged as core sites for AI compute expansion⁶, with significant projects underway in Southern Europe as well.⁷ Microsoft is planning to invest €6.7 bn in data centre projects in the Spanish region of Aragón (Butler 2024), which has led to political frictions. Advocates for the data centre expansion cite the jobs created and value added for local economies. The European Data Centre Association suggests that the data centre expansion would lead to 80,000 jobs and add €83.8 bn to GDP by 2030 (EUDCA 2025). More critical estimates point out that the actual employment effect is much lower, much of the investment spills abroad because of hardware purchases, and the land and energy implications are not given enough attention (Gabert-Doyon 2026).

Estimates suggest that the majority of new data centre real estate is mainly driven by the largest hyperscalers.

The latest and most accurate available data from datacenterHawk, from the first half of 2024 (see Figure 7), suggests that the largest buildouts still make up the bulk of the total data centre investments in the EMEA area. While this number does not distinguish between AI and conventional cloud, and also includes the Middle East, it gives a directional estimate on the prevailing investment dynamics in the sectors. This estimate likely also substantially underestimates the real hyperscaler footprint. Around 70–75 % of the built data centre capacity provided by colocation providers such as Digital Realty and Equinix is contracted by hyperscalers (Boonstra 2023). Recently, some of the hyperscalers have scaled back their contracts, creating some openings for alternatives (Donnelly 2025). However, this comes with higher lease prices to compensate for the lower credit worthiness and more uncertain demand profile.

1.3 The emergence of the alternative AI cloud market: the search for computational alternatives

Behind the shift in the macro environment, we also see a shifting composition of the market. The neocloud market emerged as a response to growing demand for specialised AI compute, known as AI accelerators. Following OpenAI’s release of ChatGPT in late 2022, a global surge in investment in generative AI research and development was accompanied by an insatiable demand for computation, particularly of high-end Nvidia chips. Hyperscale cloud providers and leading AI labs raced to secure high-end chips and data centre capacity, which was unmet by the market due to Nvidia supply chain bottlenecks in chip production, the lead time for new data centre buildouts and energy interconnection capacity issues. This demand created a structural global scarcity of advanced chips.

In response, the alternative AI compute providers, and particularly neocloud providers, have pursued revenue diversification strategies, by offering flexible contracts, specialised infrastructure configurations and cheaper compute. According to a McKinsey report, there are currently just over 100 neocloud companies globally, of varying sizes (Mazza et al. 2025). The addressable market is therefore

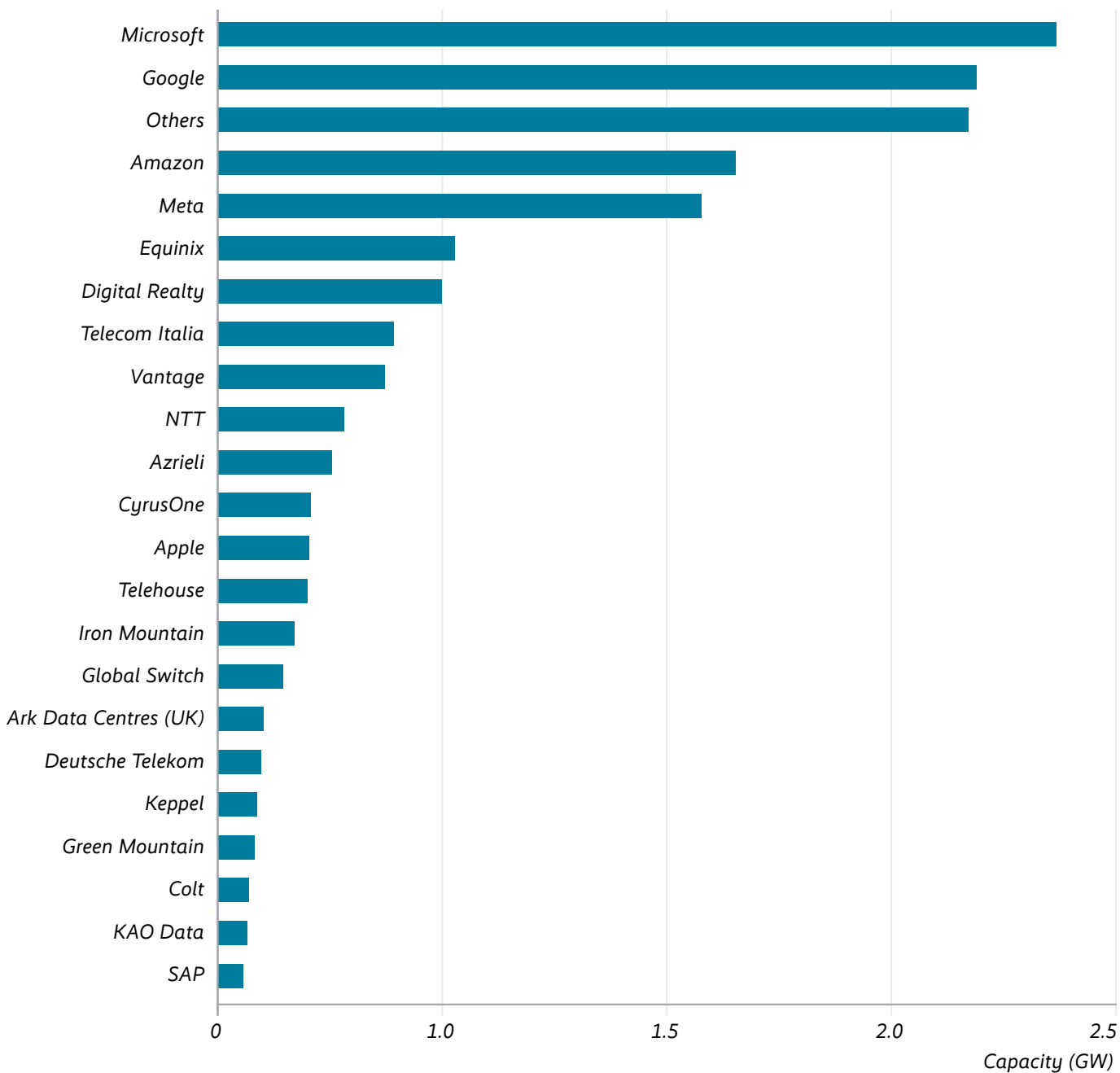
⁴ Due to interconnection queue access problems, these gigawatt scale data centres are likely to be built in a modular fashion in increments of 100 to 200 MWs. We thank Advait Arun for this notion.

⁵ One could ponder what Sam Altman’s 250 GW of power by 2033 could imply financially or in terms of energy infrastructure buildout. His vision of 1 GW per week growth, announced in his blog at <https://blog.samaltman.com/abundant-intelligence>, seems to be out of sync with the material realities of any previous of large infrastructure project. See e.g. the analysis of Gardizy & Efrati 2025.

⁶ See announcements of Coreweave at <https://www.coreweave.com/blog/coreweaves-european-expansion-lets-power-tomorrows-ai-innovations> and <https://www.datacenterforum.com/datacenter-forum/coreweave-and-bulk-infrastructure-partner-to-launch-one-of-europes-largest-ai-deployments-in-norway>; of atNorth at <https://datacentermagazine.com/news/atnorth-grows-team-for-nordic-ai-data-centre-delivery>; of Options Technology at <https://datacentermagazine.com/news/options-technology-selects-atnorth-for-cloud-expansion>; of Fluidstack at <https://insidehpc.com/2025/03/fluidstack-to-deploy-exascale-gpu-clusters-in-europe-with-nvidia-borealis-data-center-and-dell/>; of Verne and Nscale at <https://www.verneglobal.com/news/news-verne-and-nscale>; of Crusoe at <https://www.crusoe.ai/resources/newsroom/crusoe-expands-iceland-data-center-capacity>; of Verne and Nebius at <https://www.verneglobal.com/news/news-verne-strikes-10mw-deal-with-nebius-to-further-expand-europes-ai-capacity>; and of Nvidia at <https://nvidianews.nvidia.com/news/nvidia-partners-with-europe-model-builders-and-cloud-providers-to-accelerate-regions-leap-into-ai>.

⁷ See for instance activity in Spain by Microsoft at <https://www.datacenterdynamics.com/en/news/microsoft-gets-initial-approval-for-data-center-campus-in-arag%C3%B3n-spain/>; and in Portugal, there is Start Campus’s Sines Data Campus: <https://www.startcampus.pt/sines>.

Data centre building tracker in EMEA (Q1-2, 2024)



Source: Data from Omdia

relatively small, with Equinix estimating the non-hyperscaler AI compute market at €12bn by 2030 (Tairych & Delp 2025). However, they have been highlighted as a potential alternative to the entrenched dominant position of the hyperscalers.

Neocloud providers emerged with the provision of AI servers and virtual machines, or an ‘AI compute GPU-as-a-service’, offering to meet global demand for high-end computer chips. The origins of many companies lies in cryptomining, possessing parallel computing chips that became a scarce commodity after the launch of ChatGPT spurred

intense competition to develop LLMs. The crypto bubble bursting around 2019 led companies like the US cryptomining firm CoreWeave to pivot their business model towards the provision of compute for AI applications (Dave 2024), by repurposing their own GPUs and buying up GPUs from struggling cryptominers.

These neoclouds have a different cost and product profile compared with traditional hyperscale cloud providers, which offer diverse computation, infrastructure, platform and software services. Neoclouds therefore positioned themselves as alternatives to hyperscale providers. As they

do not differentiate on AI chip GPU features or product capabilities, neoclouds distinguish their products by offering cost-effective, streamlined and specialised AI compute configurations; they also cut costs by limiting their staff and infrastructure overheads. According to the Uptime Institute, the average hourly cost of an Nvidia DGX H100 instance when purchased on demand from a hyperscaler was \$98. When an approximately equivalent instance is purchased from a neocloud, the unit cost drops to \$34, a substantial 66% saving (Rogers 2025).

For the purposes of our analysis, we differentiate between two types of neocloud providers, considering market share, scale and regional presence: neocloud giants and emerging neoclouds.

Neocloud giants are leading providers by market share and scale (typically over 100k chip GPU clusters), which includes Coreweave (the largest neocloud), Crusoe and Nebius. According to SemiAnalysis, these companies are driving consolidation within the neocloud market, which is a threat to smaller providers (Patel & Nishball 2024). Many of these companies have pursued a strategy of meeting hyperscaler excess demand and relying on their baseline utilisation.

In contrast, **emerging neoclouds** offer localised and specialised AI chip GPU services. Companies like Nebul, a Netherlands based neocloud company, are smaller providers by market share and scale, and represent a longer tail of providers with less experience running data centre infrastructure, but which may represent emerging or regionally significant providers.

For the purpose of the analysis, we also include **conventional cloud**, representing regionally significant cloud providers (that are not hyperscalers) that are pivoting to offer AI compute GPUs-as-a-service. In the EU, smaller European cloud providers like OVH, Hetzner and Exoscale offer AI compute GPU products. These providers often exploit regional differentiation by aligning with domestic sovereign AI initiatives, e.g. by catering to companies that want to keep compute out of the US or China for regulatory, privacy, data security or other business reasons (Patel & Nishball 2024). The following section analyses the strategic positioning of these entities within the context of the EU's pivot towards sovereign AI infrastructure.

2. Alternative AI cloud providers and their strategies in the EU market

Europe's pivot towards sovereign AI infrastructure has created a market opening for private actors that want to capture part of the AI cloud market. This has created **unconventional market dynamics** in the European market, in which existing dependence on American hyperscalers has long made it difficult for regional alternatives to break through.

New AI-specific compute providers (called neoclouds) have emerged, along with existing cloud and telecommunications providers, which are transitioning to the AI compute space. These providers seek to strategically differentiate themselves through sovereign or specialised offerings that explicitly reduce dependencies on non-European providers. They represent **pivotal intermediaries** in achieving the

EU's political objective of alternative computational infrastructure. As such, their corporate strategies, incentives and dependencies have implications for the location, financing and control over AI cloud resources in the EU.

2.1 Market actors

We sampled 19 alternative AI cloud providers that are operating in the EU market, meaning that they have current or planned AI cloud availability within EU member states and the European Economic Area. To derive our sample, we considered global and regional market dominance, recent public announcements as well as prevalence in EU policy

Background of AI cloud neocloud providers included in the analysis (n=19)

Table 2

Provider name	Neocloud type	Company name	HQ location	European presence	Founded
CoreWeave	Neocloud Giant	CoreWeave	Global HQ in New Jersey, US; European HQ in London	Yes	2017
Crusoe	Neocloud Giant	Crusoe	Denver, US	Yes (planned)	2018
Exoscale	European Traditional Cloud	Telekom Austria	Lausanne, Switzerland	Yes	2011
Fluidstack	Emerging Neocloud	Fluidstack	London, UK	Yes	2017
GCore Labs	European Traditional Cloud	GCore Labs	Contern, Luxembourg	Yes	2014
Genesis Cloud Gmbh	Emerging Neocloud	Genesis Cloud Gmbh	Munich, Germany	Yes	2018
Hetzner	European Traditional Cloud	Hetzner Online GmbH	Gunzenhausen, Germany	Yes	2018
IONOS	European Traditional Cloud	IONOS	Montabaur, Germany and Philadelphia, US	Yes	1988*
Mistral Compute	Emerging Neocloud	Mistral AI	Paris, France	Yes	2025
Nebius	Neocloud Giant	Nebius Group	Netherlands	Yes	2024**
Nebul	Emerging Neocloud	Nebul	Amsterdam, Netherlands	Yes	2022
NScale	Emerging Neocloud	Nscale	London, UK	Yes	2023
OVHcloud	European Traditional Cloud	OVH Groupe SA	Roubaix, France	Yes	1999
Scaleway	Emerging Neocloud	Illiad Group (parent company)	Paris, France	Yes	1999
Sesterce	Emerging Neocloud	Sesterce	Marseille, France	Yes	2018
Taiga Cloud	Emerging Neocloud	Northern Data (being bought by Rumble) (DCD, 2025)	Dublin, Ireland	Yes	2023
UpCloud	European Traditional Cloud	UpCloud	Helsinki, Finland	Yes	2011
Vast.ai	Emerging Neocloud	Vast.ai	Los Angeles, US	Yes	2018
Verda ***	Emerging Neocloud	Verda	Helsinki, Finland	Yes	2020

* Became IONOS in 2020, splitting from the webhosting part of the company

** Nebius split from Russian firm Yandex in 2024.

*** formerly Datacrunch

conversations about AI compute. Finally, **we sense-checked our selection with experts in AI and cloud policy and development**. Our data set covers both neocloud companies (n=13) and conventional cloud companies (n=6) providers, and considers market share, market presence and political salience (e.g. whether they are involved in digital sovereignty partnerships).⁸ In Table 2 below, you will find an overview of AI cloud companies included in our analysis.⁹ As this is exploratory research based on publicly available data on 19 providers in a rapidly evolving and highly opaque market, the findings should not be read as comprehensive summary of the European AI cloud landscape, but rather as a sample to illuminate broader patterns in this market.

All 19 sampled countries have a current or planned EU presence in terms of AI cloud availability. Across the 13 neocloud companies included, 8 are headquartered in the EU (Genesis Cloud, Verda, Scaleway, Sesterce, Mistral Compute, Nebul, Taiga Cloud and Nebius), while 5 are headquartered outside the European Union (Vast.ai, NScale, Fluidstack, CoreWeave and Crusoe).¹⁰ These include the world's largest neocloud providers by global market share: CoreWeave, Crusoe (which are US companies), and Nebius (which is a Dutch company, formerly European branch of the Russian cloud provider Yandex). The remaining 6 companies fall into the category of conventional cloud providers, which have their headquarters in Europe (UpCloud, Exoscale, Hetzner, OVHcloud, GCore Labs and IONOS).

Additionally, **a significant proportion of Europe-operating neoclouds** appear to be pursuing differentiation strategies through a geographic shift towards the Nordics. Of the 19 companies surveyed, 14 companies have cloud capacity in the Nordics with Finland (n=8), Iceland (n=4), Norway (n=8) and Sweden (n=6).

2.2 Market strategies

An analysis of these providers' market strategies, funding and product offerings reveals four approaches used to carve out a sustainable and regionally significant niche (see Table 3):

- **Specialisation:** sovereign AI, niche markets and sustainability
- **Upgrading:** start-ups/SMEs, full-stack
- **Consolidation:** corporations, sovereigns
- **Market place:** platform, aggregators

Rather than a set of fixed strategies, these are often combined by specific neoclouds across various gradients based on their company objectives.

Alternative AI cloud companies pursuing a market strategy of **specialisation** are carving out niche markets outside of US hyperscale cloud providers. Such niche markets include a focus on sovereign AI capabilities; companies like Verda, Sesterce and Nebul emphasise European sovereign cloud and divestment from US infrastructure, through data centres located in European jurisdictions, GDPR-compliant architectures and full-stack control. For example, Sesterce positions itself as the 'third way between hyperscalers and inaction', explicitly arguing that it is building 'the backbone of European AI independence'¹¹, with a €52bn investment project to build a total of 1.5 GW of sovereign computational power.¹² The **primary target** for specialisation companies are sovereign and highly regulated industries, on the premise that these sectors will absorb the friction and cost associated with migrating from hyperscalers. Firms with a presence in the Nordics (Genesis, Verda, etc.) are also focusing on sustainability as a competitive advantage, with an emphasis on environmentally friendly data centre locations.

Upgrading strategies refer to efforts to move away from the provision of bare computers to more high value-added services. First, companies seek to explicitly target SMEs and start-ups to create long-term customer relationships. Another strategy is to move up the value chain to provide full-stack services. One key provider is the French company Scaleway, which provides computational capacity for several French-headquartered AI companies, such as Kyutai, Mistral AI and H Company. Scaleway provides full-scale AI solutions and seeks to avoid becoming solely used for computational capacity. New entrants are also emerging in this field: in June 2025, Mistral AI, the most prominent large-scale AI developer in Europe, announced it was pivoting to a cloud infrastructure company.¹³ Under the label of 'Mistral Compute', the company offers customers 'GPUs, orchestration, APIs, products and services in whatever form factor they need', expanding from 'bare-metal servers to fully managed PaaS.' Partnering with selected European enterprise customers, the company teamed up with UK-headquartered Fluidstack to expand a 100 MW data centre in Essonne, France (Business Wire 2025), although, as per Bloomberg reporting, Fluidstack is backtracking on this investment (Berthelot 2026). Mistral AI also announced plans for a separate data centre in Paris with over 1.4 GW of capacity, in partnership with the

⁸ Variables included in the dataset (collected in November 2025) are as follows: Provider, Location (HQ), Date Founded (Year), Net Worth (Revenue), Geographic Presence of Data Centres (Location, Number and MW), Chip Types and Capacity (Name and Number), Product Offerings (Description), Corporate-State Collaborations (Description), Known Customers (Description) and Business Strategy (Description).

⁹ Our analysis is limited by data availability. As there is limited market data on the EU neocloud market analysis, our data collection is often descriptive rather than standardised. Where the AI cloud segment of a company is unavailable, we infer data from the total cloud share.

¹⁰ In contrast to previous sections' broader focus on AI cloud in Europe, for this section's analysis of neoclouds we explicitly distinguish between the neoclouds that are headquartered in the EU and those that are not. This is relevant, because the next sections will examine to what extent neoclouds can help advance the EU's ambitions for AI sovereignty.

¹¹ <https://www.sesterce.com/blog/a-third-path-for-ai-sovereignty-beyond-hyperscalers-and-inaction> (last accessed on 23 April 2026).

¹² <https://www.sesterce.com/blog/building-france-s-ai-backbone-by-2030> (last accessed on 23 April 2026).

¹³ <https://mistral.ai/news/mistral-compute> (last accessed on 23 April 2026).

Alternative AI cloud market strategies

Approach	Strategy	Illustrative examples
Specialisation Carving out niche markets to differentiate from and complement hyperscaler offerings.	Specialisation (sovereign AI) → focussing on AI cloud provision to markets outside the US or China → Often backed by government or regional champions that value independence from hyperscalers (based on regulatory, privacy, data security or geopolitical concerns)	Most of the EU providers
	Specialisation (niche markets) → focussing on specialised and regulated verticals e.g. healthcare	Genesis Cloud
	Specialisation (sustainability) → focussing on sustainable infrastructure, often exploiting cooler climate countries e.g. Nordics	Genesis Cloud, Verda
Upgrading Moving up the value chain; focussing on AI start-ups as their core customer profile, rather than chasing large enterprises; maybe expanding to AI-native enterprises.	Upgrading (start-ups/SMEs) → focussing on start-ups and SMEs requiring smaller systems	IONOS, Exoscale, Hetzner,
	Upgrading (full-stack) → capturing start-up/SME market and upgrade to AI-native enterprises through vertically integrated solutions	Scaleway, Mistral Compute, Vast.AI (as a software partner for other companies)
Consolidation Getting absorbed by hyperscalers, telcos or sovereign buyers	Consolidation (corporations)	Nscale
	Consolidation (sovereigns)	Verda
Marketplace Aggregating demand and supply; carried out by capital-light companies.*	Marketplace (platform) → Companies that match AI compute owners to buyers. May also provide IaaS infrastructure, but don't have expertise in deploying compute	Vast.ai, Compute Exchange
	Marketplace (aggregators) → Companies that create a marketplace for AI compute owners to offer compute to buyers	Nvidia (partnerships and the DGX Cloud Lepton program). Together.AI

* See Patel & Nishball 2024

UAE sovereign tech fund MGX, which would be the largest AI cloud available in Europe (Swinhoe 2025). Finally, it recently announced having raised \$830 million in debt to finance a new 44 MW data centre near Paris (Mukherjee & Marchandon 2026).

Consolidation refers to increasing collaboration with existing industrial companies and sovereign compute projects. For example, NScale announced a strategic partnership with the Finnish networking giant Nokia to develop AI-ready computational infrastructures, with Nokia also participating in the funding round with the

company.¹⁴ Similarly, Stargate Norway, a joint venture that aims to deliver 100,000 Nvidia GPUs by the end of 2026 to OpenAI's first data centre in Europe, will be delivered by Nscale, in close conjunction with the Norwegian investment giant Aker, which also took a 9.3% stake in the company.¹⁵ This consolidation also has a sovereign dimension, with providers supporting the bids for Gigafactories. One example is Verda's role in the Baltic-Polish Gigafactory bid for one of the €5 bn projects on offer, overseeing the AI Gigafactory's technical architecture, operational planning, deployment and life-cycle maintenance.¹⁶

¹⁴ <https://www.nokia.com/newsroom/nokia-and-nscale-partner-to-accelerate-ai-infrastructure-build-out/> (last accessed on 23 April 2026).

¹⁵ <https://news.cision.com/aker-asa/r/aker-finalizes-investment-in-nscale-and-joint-venture-in-narvik,c4246455> (last accessed on 23 April 2026). To note, Stargate Norway is a distinct joint venture between Nscale, Aker and OpenAI. It should not be confused with the proposed U.S.-based Stargate supercomputer project between Microsoft and OpenAI; for more see Craske, 2025.

¹⁶ <https://verda.com/blog/datacrunch-latvian-alliance-submits-proposal-for-eus-ai-gigafactory> (last accessed on 23 April 2026).

Another trend involves nascent aggregation models; specialist capital-light platforms are experimenting with matching fragmented supply to niche demand. While the key player in this trend is Nvidia's **DGX Cloud Lepton** (see below), capital-light providers like **Vast.ai** and **Together.AI** are also beginning to aggregate fragmented supply for smaller buyers. They seek to match the centralised demand to fragmented non-committed supply in the form of a market for GPUs. One example is the Compute Exchange, which functions as an on-demand marketplace for GPU compute. In their submission to the European Commission's call for evidence for an impact assessment on 'AI Continent – new cloud and AI development act', they noted that their platform would solve the 'cost, capacity and concentration challenges that EU innovators face in the current inefficient compute market.'¹⁷

In summary, there is an emerging set of actors that position themselves as alternatives to the existing hyperscalers and as partners to the EU's sovereignty push. Some of these European providers offer much more competitive prices than the legacy behemoths, providing leaner compute stacks without a bundle of extraneous software, with low latency in inference being highlighted as an additional competitive advantage (Dahlstroem & Stephens 2025). While not the core focus of this study, telecommunications operators, such as Telenor in Norway¹⁸ and Deutsche Telekom in Germany (Kaesebier, Rugamer & Marchandon 2025) are also pursuing exposure to the AI cloud market (Shrivastava et al. 2025), with a role in some of the leaked Gigafactory bids (Serrano 2025). These actors are also politically active stakeholders, as shown in the responses to the call for evidence for on the upcoming EU Cloud and AI Development Act (CAIDA).

¹⁷ Compute Exchange submission to the call for evidence on an impact assessment on AI continent – new cloud and AI development act, 1.7.2025 (reference F3571381), https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/14628-AI-Continent-new-cloud-and-AI-development-act/F3571381_en.

¹⁸ <https://www.telenoraifactory.no/> (last accessed on 23 April 2026).

3. Challenges for the EU AI Cloud

The development of the alternative AI cloud ecosystem faces structural challenges that may subvert the intent of the EU’s policy towards more technological sovereignty.

3.1 Physical constraints and bottlenecks

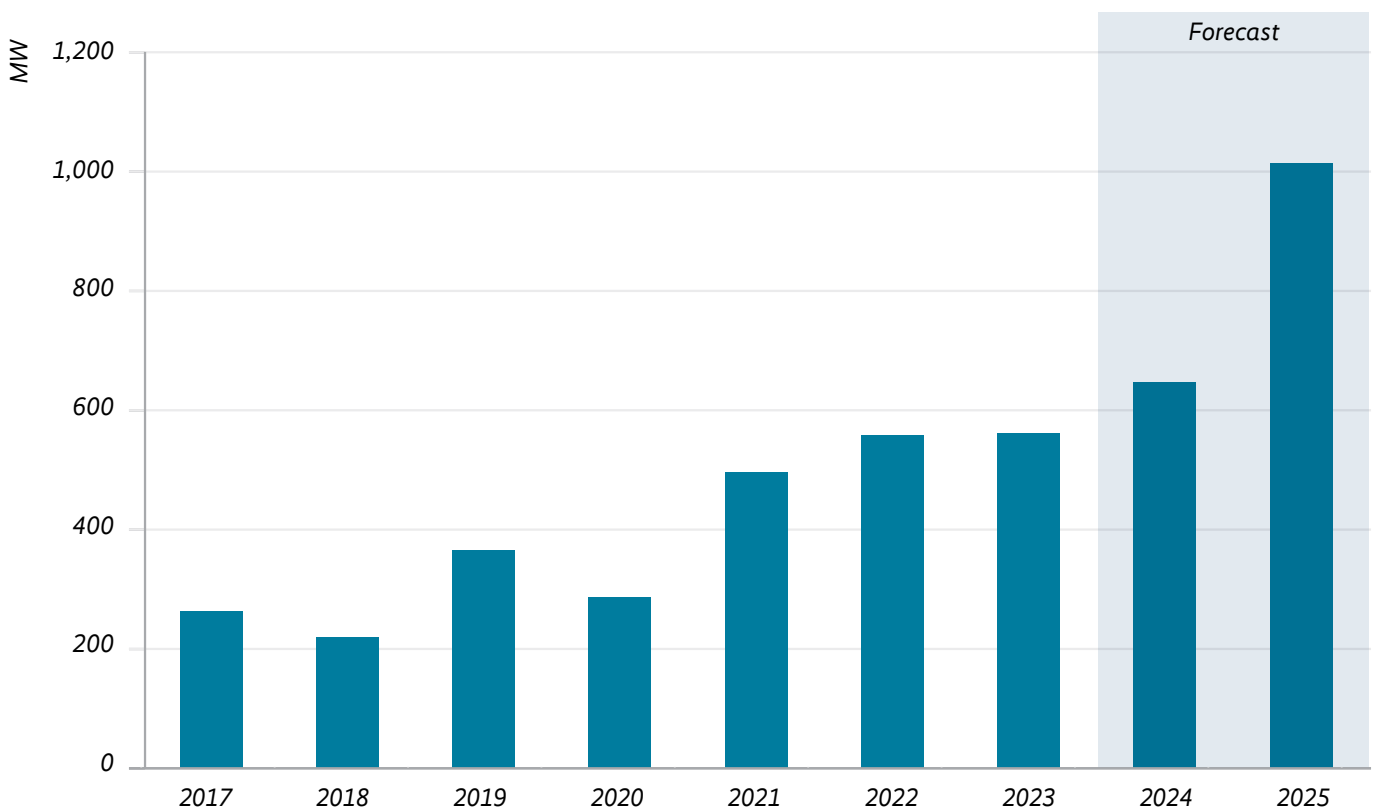
Physical constraints regarding land use, power generation and grid transmission are out of sync with the aspirations of data centre growth. The EU’s objective of ‘tripling data centre capacity’ is primarily limited by these infrastructural material realities rather than just directly AI-related capital or hardware availability. AI computing requires vastly more power than previous cloud technologies, creating major challenges for electricity grids and transmission infrastructure. Building energy infrastructure is typically a long-term process,

with delays in obtaining permits and procuring the necessary energy components, which are currently highly sought after.

Beyond environmental questions and the capital expenses required, there are distributional implications (Leppert 2025). For example, largely as a result of data centre-driven buildout in Ireland, the Commission for Regulation of Utilities (CRU) projects an 8 to 21% increase in energy prices for households over the next five years, while the data centres and other large energy users will see a 14% decline (Smyth 2025). While some European companies are positioning themselves to capture markets in the cooling and heat re-use solutions, data centre buildouts of the scale proposed is likely to lead to challenging political trade-offs and tensions between other policy objectives, such as energy security and the climate transition.

European data centre supply projections

Figure 8



Source: CBRE 2025d.

3.2 Demand concentration and the market risk

The demand structure of the AI market reinforces dependence on the incumbent players. The key 'demand sink' of current AI compute providers has been hyperscalers themselves and some of the leading AI labs that provide high-volume, long-term demand for computation capacity.¹⁹ This is reflected in the revenue contributions of hyperscalers to neoclouds. For example, 62–77% of CoreWeave's \$5.25bn revenue is made up of Microsoft contracts and providing AI compute for OpenAI (Zitron 2025). Across the board, removing Microsoft, Amazon, Google, OpenAI and Nvidia's contribution to CoreWeave, Nebius and Lambda Cloud's revenue leaves only \$1bn in combined revenue (Zitron 2025). Other sources of demand are often comparatively more uncertain and fragmented to provide sufficient certainty for future cash flows.²⁰ This parallels broader trends in the AI economy, where any paths to profitability are closely tied to large American tech companies. This binds the nascent neocloud market to the demand and incentive ecosystem that surrounds the hyperscalers.

The rapid depreciation of AI chips exacerbates the demand problem (Martindale 2025). High hardware costs and related capital expenses for infrastructure require a consistently **high utilisation rate** to recoup the investment. Beyond the initial procurement of chips, the pressure to keep pace with rapid refresh cycles shortens the window in which profit must be generated, driving up the required utilisation rate even further.²¹

For instance, if a \$40,000 GPU becomes technically obsolete in just two years, a provider must ensure it is running billable workloads nearly 24/7 just to break even before the next generation arrives. This has led to the prevalence of long-term 'take-or-pay' contracts to ensure predictable cash flows; that, in turn, bolsters the financial dominance of existing hyperscalers, as they are one of the few parties that can agree to guarantee 24/7 demand for workloads several years in advance. The volume of demand from regulated industries – such as defence, finance and healthcare – that form the primary target market for **sovereign specialised providers** is unlikely to be sufficient to amortise these hardware costs in a rapid obsolescence cycle, creating a systemic competitive disadvantage.

In addition, the financing mechanics create sustained scale benefits that favour incumbent players in the global AI cloud market. While hyperscalers have largely funded their AI infrastructure buildout from cashflows, neo-

cloud providers fund their investments through asset-backed debt and project finance (Lopatto 2025). This means that they often have high interest rates to contend with. The financing structure makes the alternative neoclouds more sensitive to demand fluctuations. If the AI cycle starts slowing down, the demand fails to materialise or the neoclouds are unable to capture value, alternative AI cloud providers are likely to experience cash-flow issues. This is particularly concerning given the unproven profitability of genAI applications that are driving demand for this infrastructure (Brennan, Kak & Myers West 2025). This is unlike hyperscalers, which will be able to balance demand fluctuations across their other, profitable business segments and absorb cyclical changes in their balance sheets. This gives them the potential to scoop up the competition when the market consolidates.

3.3 Supplier concentration and the Nvidia risk

Against these structural pressures, Nvidia is propping up this ecosystem, posing a concentration risk. In order to diversify demand sources for its highly sought-after AI chips away from hyperscalers and encourage competition in the ecosystem, Nvidia has been eager to partner with and cultivate alternative ecosystems. As such, Nvidia has indirectly become a geopolitically salient partner for European parties with technological sovereignty aspirations. Nvidia has forged numerous strategic partnerships with actors in our ecosystem, with 385 partners in the EMEA market. In our analysis, based on data collected in November 2025, we noted that **~78% of the alternative AI cloud providers have strategic partnerships with Nvidia**, securing preferential access to the chips and, in some cases, receiving equity investments from the company. This entrenches demand for Nvidia chips that is enforced in several ways: providers like CoreWeave are financed by Nvidia, enabling them to purchase Nvidia chips, which have since been used as collateral to purchase more Nvidia chips (Lopatto 2023). This aligns with previous research on the contemporary importance of Nvidia in the tech economy as a key background facilitator and ultimate beneficiary of the latest epoch of AI buildout.

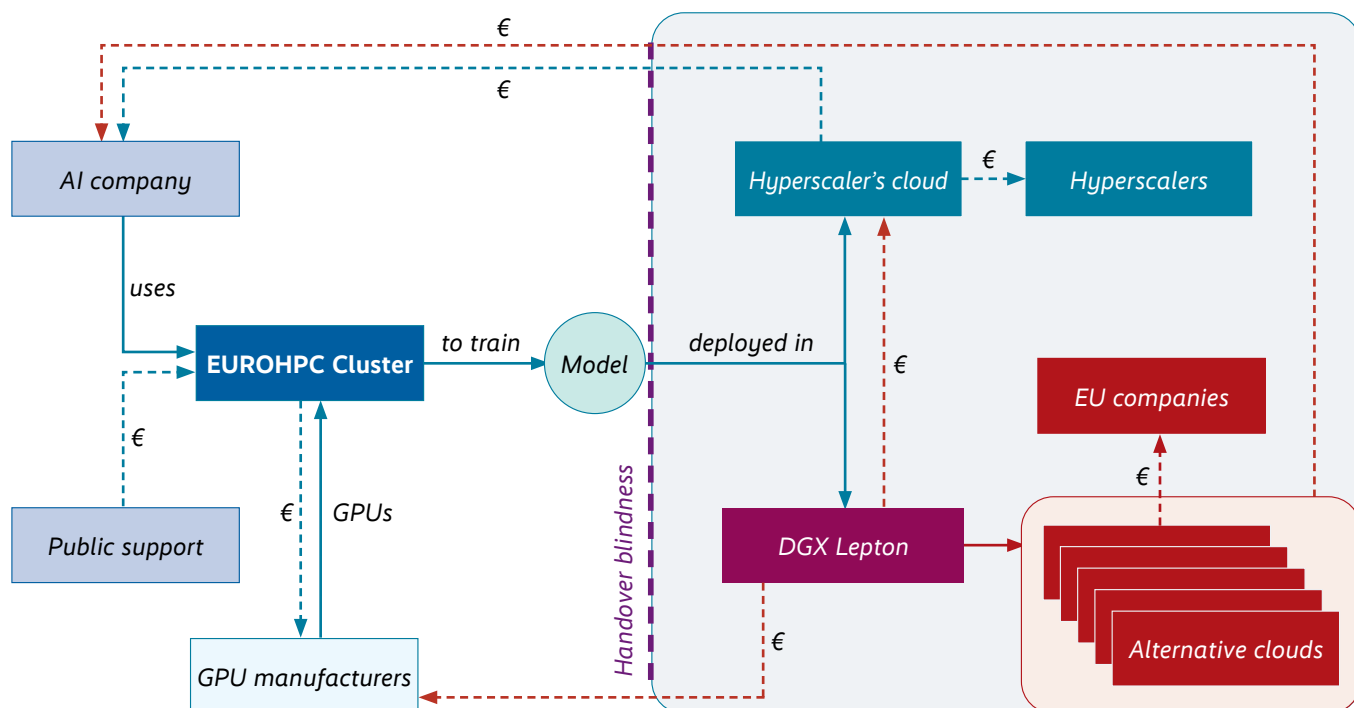
Nvidia has been embedding its current market position in several ways, for instance by: investing in leading European AI labs (Iyer 2025; Gerlat 2025; Thole 2025); partnering with European VC funds like Accel, Elaia, Partech and Sofinnova Partners (Nvidia 2025), as well as top European companies such as Nokia (Nokia 2025), Deutsche Telekom (Deutsche

¹⁹ As to the potential lack of demand for the Gigafactories in the EU, see Hess & Sieker 2025, and Renda & Kyosovska 2025.

²⁰ Note here that, in some cases, the larger companies will contractually agree to offtake capacity from the neoclouds. For instance, Nvidia has agreed to buy off all the unused compute capacity from Coreweave as a way to generate stable demand (Lopatto 2025). This is also a way to compete on the hardware side. For instance, we believe Google's provision of offtake agreement to Fluidstack has to do with the effort to convince the company to use Google-made TPUs (Huang, Clark & Jin 2026). The offtake agreements become the implicit backstops for neoclouds, pointing to the importance of long-term contracts in structuring the incentives of different actors.

²¹ Many of the GPU providers argue that trailing edge chips have a long life cycle, as they can be used for less computationally cutting edge operations. For a discussion of those claims, see Butler 2025.

How Nvidia may capture the alternative AI compute market through aggregator strategy



Source: Authors, based on Arun 2025

Telekom 2025) and Siemens (Nvidia 2026), to generate demand; backstopping new European AI start-ups²²; and partnering and investing with asset managers to drive the adoption of Nvidia-specific architectures for new data centre buildouts (Butler 2026; Brookfield 2025b). It is only a slight exaggeration to note that the pursuit of EU AI sovereignty is largely shifting the nature of existing dependencies from established American cloud service providers to the chip giant.

This concentration and entrenched market position is a response to the tight competition in chip markets. In late 2025, the capacity gap of the Nvidia chips seemed to be narrowing, with world-leading LLMs being trained with alternative chips, such as Google's own TPUs (Bradshaw 2025). Other hyperscalers, such as Amazon, are also advancing in the development of their in-house chips, providing them for use by some leading AI labs (Clark 2025). AMD, a leading alternative chip designer, is also catching up with its newest chips (Chen et al. 2025). By integrating itself into the infrastructural, computational fabric of European sovereign AI, facilitating partner ecosystems (Siemens 2026), diversifying markets and *de facto* buying prospective com-

petitors (Groq 2025; McKenna 2025), Nvidia is reinvesting profits into strategies that secure a sustainable market position for itself and technical ramp that locks in revenue if and when the current large-scale, LLM-based AI cycle will slow down. This protects the company from existing competitors and limits the chances of alternative chip designers breaking into the market.²³

Dependence on Nvidia comes with strategic risks for alternative AI cloud providers, specifically neoclouds.

First, the interests of Nvidia's product roadmap and neoclouds' business models are not aligned. Nvidia seeks to launch a new product every 12–18 months to sustain revenue growth. For neoclouds, this creates a conundrum, as the value of their existing chips will depreciate with accelerated schedules that come with new product launches. This has led to tensions in the market, with neoclouds struggling to guarantee enough demand for their chips to recoup their investment (Butler 2025).²⁴

Second, Nvidia is increasingly capturing the market of alternative providers. Providing 'bare compute' leads to a 'commodity trap', which is highly contingent on the supply

²² One notable example is Domyn in Italy, building a 'sovereign AI cluster'. See <https://www.domyn.com/compute> and <https://www.youtube.com/watch?v=bnVwzjtkc04&feature=youtu.be>.

²³ See, for instance, Axelera AI, which has received substantial direct equity funding in a €63m Series B from the European Innovation Council and the Dutch National Development Bank Invest-NL; see https://eic.ec.europa.eu/news/first-companies-put-forward-major-investments-under-eic-step-scale-scheme-2025-04-03_en and <https://www.invest-nl.nl/impact/ons-portfolio/axelera-ai> (last accessed on 23 April 2026).

²⁴ The new leasing arrangements could be a Nvidia response to manage the outstanding stock of old GPUs, see Efrati 2025.

crunch of the AI chips that followed the launch of LLMs like ChatGPT. On 18 May 2025, Nvidia launched **DGX Cloud Lepton**, a unified interface for accessing pooled GPU compute from various providers.²⁵ While this pools demand and offers access to blue-chip customers, it also erodes differentiation. This erodes margins and prevents the development of high-level, long-term business models. Although some critics focus on technical flaws (SemiAnalysis 2025), the real threat is the dumb-piping of competitors²⁶, effectively turning them into infrastructure utilities and blocking their path to higher value-added services. Although this is efficient for the market, this erodes the margins required to amortise expensive hardware.

3.4 Market structure challenges

The niche markets are also suffering from oversaturation. In order to avoid directly competing with hyperscalers, European alternative AI cloud providers often target the same market segment of sovereignty-minded customers in regulated enterprises that prefer European-hosted computing.²⁷ The demand in these AI cloud markets is substantially smaller than the broader AI compute market, with Equinix estimating the size of the non-hyperscaler AI-compute-as-a-

service market at €12bn euros by 2030 (Tairyach & Delp). This submarket is crowded in Europe, with several providers sharing the same tactic for competing in this space, pushing down margins. Lastly, the public AI factories and public-private Gigafactories aim to provide compute essentially to this same segment, further saturating the finite demand.

Alternative AI cloud providers are having to navigate an ecosystem where the incentives afforded to other players are structurally stacked against them. The AI-compute value chain consists of chip manufacturers, colocation providers, compute providers, credit providers, end users and a plethora of alternative actors that provide financing, cooling solutions, energy and connectivity networks in this ecosystem. These complex relationships and incentives create constraints in the alternative computation ecosystem. Colocation providers and data centre operators prefer hyperscalers as customers: based on our analysis of data-centerHawk data, **hyperscalers enjoy on average an estimated ~25% discount over wholesale customers and ~50% over retail customers on the colocation prices in the European Union because of their role as preferred, stable long-term customers** (see Figure 10). End users, such as start-ups, prefer to use hyperscaler's clouds because of the ubiquitous free compute credit programmes, the

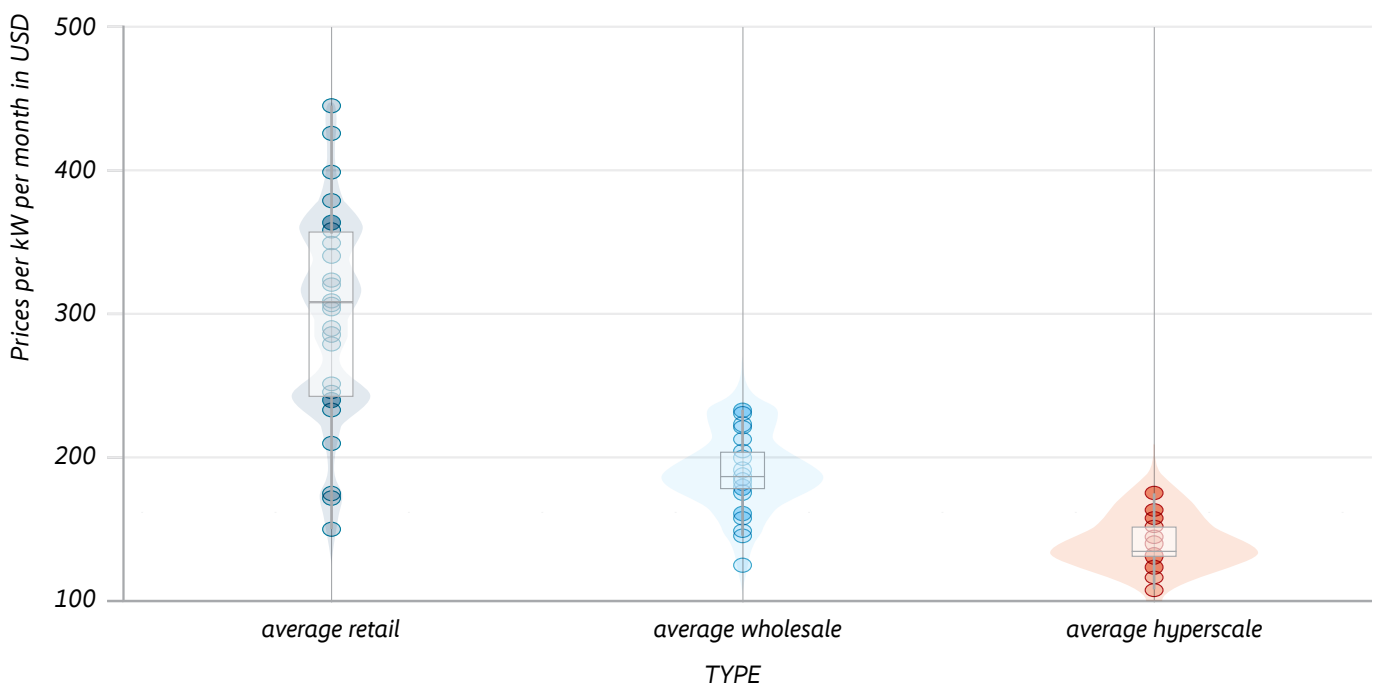
²⁵ The implication is also that this further centralises the market due to the software lock-in; pooling requires interoperability on the hardware level, which in practice will be challenging to achieve across different manufacturers. This might strengthen the existing position of the leaders.

²⁶ https://en.wikipedia.org/wiki/Dumb_pipe.

²⁷ For the parallel strategies of European AI firms active in the model layer, see Saari 2024.

Average colocation pricing per customer type in European data centre locations

Figure 10



Source: Data from datacenterHawk

existing engineering skill base tied to these platforms and convenient access to existing AI models via their APIs (Kedrosky 2024). Private credit providers and asset managers that want exposure to the AI data centre buildout prefer hyperscalers because of their investment grade status, stable demand and strong cash flows. These structural incentives together create a gravitational pull towards the AI compute market that affects policy implementation. This can lead to counterintuitive policy outcomes, such as the case of a recent European Investment Fund (EIF) data centre investment described below.

3.5 Industrial policy for hyperscalers? The case of EIF, the CDP and Apto's Lacchiarella site

To illustrate the structural incentives that favour hyperscalers in arms-length industrial policy interventions, consider recent efforts by the European Investment Fund (EIF) and the Italian national development bank Cassa di Risparmio di Milano (CDP) to support European technological independence.

On 12 May 2025, CDP and EIF jointly announced a €200 million cornerstone, catalytic investment to Allianz SE subsidiary PIMCO's European Data Centre Opportunities (EDCO) Fund,

which was set up to support the development of European data centre infrastructure.²⁸ The data centres will be built by PIMCO's fully owned subsidiary Apto.

Just two days later, Apto announced a new project in Lacchiarella, near Milan.²⁹ The project was announced as the largest data centre campus in Italy, with an estimated area of 228,000 m², 300 MW of power consumption and an estimated €3.4 bn of total planned investment. Nominally, this seems like a substantial investment to support the development of digital sovereignty in the EU. Indeed, in the press release of the investment, the CEO of CDP noted that:

“investment in the PIMCO European Data Centre Opportunity Fund reflects CDP Equity's strong commitment to accelerating digital transformation and strengthening strategic infrastructure in Italy and across Europe. By supporting data centre development, we are helping to lay the foundations for a more competitive, innovative and resilient digital economy”,

with the Chief Executive of the EIF echoing the sentiment, noting that the

²⁸ See EIF's announcement at: <https://www.eif.org/press/all/eif-and-cdp-equity-jointly-invest-eur200-million-in-the-pimco-european-data-centre-opportunity-fund-to-expand-digital-infrastructure-across-europe>. (last accessed on 23 April).

²⁹ See Apto's announcement at: <https://aptodc.com/apto-announces-plans-for-italys-largest-data-centre-campus/>. (last accessed on 23 April).



Presumptive location of the site in the fields of Lacchiarella

“investment underscores our commitment to fostering a strong digital ecosystem in Europe. (...) By supporting the development of data centres, we are not only addressing the immediate demand for data infrastructure, but also laying the groundwork for sustainable growth and innovation in the digital economy.”

However, the market incentives, as outlined above, limit what this approach to pursuing strategic sovereignty can achieve. Without any controls, there is a possibility that, instead of challenging hyperscaler dominance and creating alternative infrastructure, these kinds of interventions unintentionally end up supporting American hyperscalers. In other words, there is a possibility that public capital will be used to finance the buildout of computational infrastructure that will be exploited by US hyperscalers.

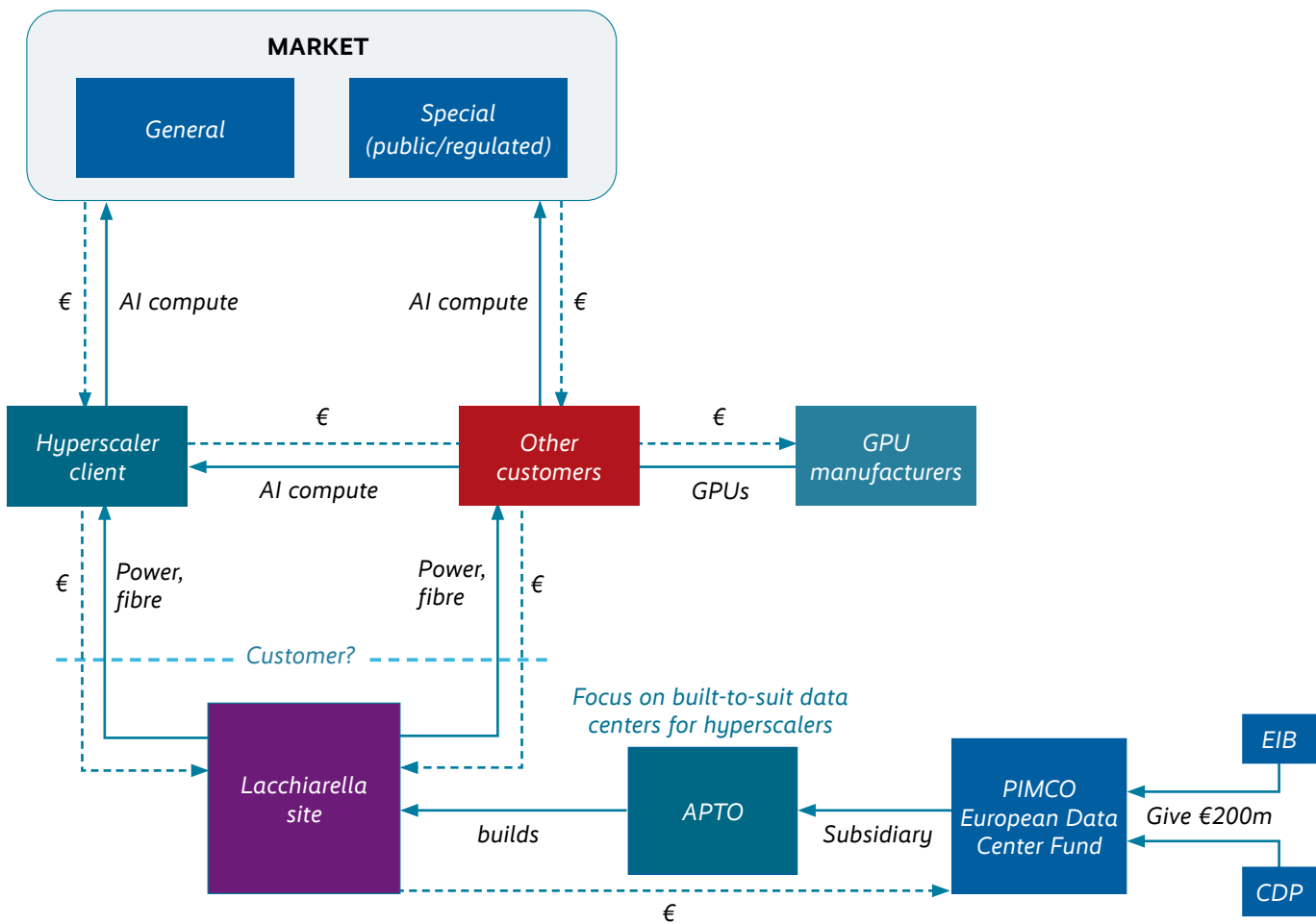
The project does not disclose the intended end user for the site. EIF, CDP and Apto did not respond to our requests for further information on the end user, the connection between the €200 million and the commencement of Lacchiarella project, and other details regarding the project financing.³⁰

In the absence of public disclosure, we analysed the structural incentives that determine plausible outcomes. First, for a €3.4 bn and 300 MW AI data centre to qualify for a data centre fund, there often needs to be a designated pre-leased anchor tenant contractually bound to the project. We are not aware of any European company that has the scale or capacity to engage in such an offtake agreement. Even pooled and designated demand, such as the Deutsche Telekom Manufacturing AI Cloud, is three times smaller in investment size. From the data centre builder perspective, a need for a stable and long term customer makes an American hyperscaler a highly favoured candidate as the end user, compared with riskier, short-term leases by multiple

³⁰ After the writing process for this paper ended, we received limited information from EIF on the project. See Saari, L. & Kaltheuner, F. (2026) Issue #16 EU AI Cloud – Industrial Policy for the Hyperscalers? Case Apto and EIF. EU AI Industrial Policy Monitor, at <https://euaipolicymonitor.substack.com/p/issue-16-eu-ai-cloud-industrial-policy>

Figure 11

The example of APTO



Source: Authors, based on Arun 2025

tenants. This is also driven by the financing conditions imposed by the data centre providers: by having investment-grade tenants in their data centres, this limits their own risk in financial markets, which helps them to secure lower interest rates on their debt.

Second, the data centre builder working on the site, Apto, states on its website that it specialises in manufacturing ‘built-to-suit’ data centres for hyperscalers.³¹ This means custom-built data centres that accommodate the specifics of large technology corporations. This further supports the hypothesis of a large hyperscaler end user, as such custom-built infrastructures are not easily fungible or readily repurposable for use by alternative customers. By locking in the infrastructure to the specifications of the large hyperscalers, the infrastructure is reserved for their capacity.

There are three possible counterpoints to our argument: first, the EIF and CDP cornerstone investment and the almost instantaneous commencement of the work in the Lacchiarella site could be a coincidence. Second, PIMCO’s EDCO could be able and willing to invest in a €3.4 bn data centre project without a prospective client in mind, hoping to secure demand as the data centre will be ready, in a wholesale or retail business model. Third, it is also possible that Apto, instead of tailoring its existing process knowledge to hyperscaler customers, uses general modular blueprints for the site to ensure that a multi-tenant strategy remains feasible. Without further information, we cannot conclusively confirm these theories to be true or false; however, we find these scenarios to be unlikely.

For this analysis, it is enough to highlight the challenges that horizontal industrial policy, such as fund portfolio investments, faces when developing alternative infrastructures in the captured AI market. The strong pull of the computational market, combined with the incentives for stakeholders, reinforces structural advantages for existing hyperscaler infrastructure. Given the enormous capital requirements and high levels of risk, the safest – and often most profitable – strategy is to integrate with existing dominant players rather than compete against them.

While the EIF’s InvestEU mandate is linked to the development of European strategic technologies, the realities of data centre financing tend to steer these objectives toward existing hyperscalers. In effect, this shifts the meaning of ‘strategic’ technology away from genuine infrastructural sovereignty toward less pressing issues, such as data residency. While nominally in line with the strategic objectives by creating a data centre infrastructure that is European owned (Apto) and European domiciled (Lacchiarella, Italy), the *de facto* service dependency still remains with the US hyperscalers.

Whether this is a problem is subject to perspective. For the (partly European) financial intermediaries, the data centre

development project is a lucrative opportunity, with a stable and predictable cash flow resulting from the exposure to the AI investment supercycle, driven by hyperscalers. From the perspective of the EIF balance sheet, the stable and secure rental income generated from the hyperscaler can be used to finance other important initiatives.

This portfolio-investor mindset suggests there is a mismatch between the realities of data centre capitalism and the policy objectives of supporting alternative European infrastructures: if the only data centre infrastructure investment on the EIF balance sheet in 2025 turns out to support an American hyperscaler buildout in Italy, the challenges for developing alternative European computational infrastructure are apparent.

³¹ See <https://aptodc.com/>. (last accessed on 23 April 2026).

4.

Policy implications: questions for EU AI policy

The findings³² put forward in this report **highlight the challenges** in implementing the planned European cloud buildout (European Commission 2026) that will have to be considered and addressed. The market incentives available to actors in the AI cloud ecosystem are shaped around the hyperscalers, making the latter the likely beneficiaries of horizontal policy interventions. The findings also raise several guiding questions in preparation for policy debates: what kind of cloud are we talking about? Where will it be hosted? What are the infrastructural requirements? Who will be the end users? And what use cases will there be?

The findings also **reveal some tricky political tradeoffs**. Nvidia has been identified as the inherent facilitator of EU AI sovereignty. It bears noting that Nvidia's own strategic business interests are not always aligned with the European public interest. One especially crucial point relates to Nvidia's product roadmap. While updating hardware every 12-18 months and racing to build ever larger computing clusters is good for the company's shareholders and its strategic objectives, this does not obviously benefit the European tax payer. The public sector ends up bankrolling Nvidia's product roadmap by procuring chips for the EU's Gigafactory buildout.

Industrial policy is often justified on the grounds that it solves demand bottlenecks and provides predictability to domestic providers. But in this case, it also implicitly commits the EU to a specific paradigm of AI buildout that locks in particular technological choices.³³ The increase in supply also glosses over the most crucial problem: the lack of aggregate compute demand outside hyperscalers and few leading AI labs (Hess & Sieker 2025). If the public sector takes as its task to provide that anchor demand, this might lead to the forced adoption of AI technologies to provide support for the local companies instead of crafting a sovereign technology policy in the public interest.

In the case of Gigafactories, the obsolescence risk is also shifted to the European taxpayer. Asset manager Brookfield predicts that the H100 and H200 chips will be phased out by 2028. The EU Gigafactories that have been proposed are now aiming to rely on 100,000s of H100 chips, with the

first operational centres built in 2027-2028 (Brookfield 2025a, p.11). This points to a fundamental mistiming and potential strategic miscalculation at the heart of the European data centre buildout.

Policy considerations

Policy should prioritise energy and grid investments before the acquisition of hardware. As the most durable element of AI infrastructure, power generation and transmission provide a 'no-regret' foundation; even if AI demand fluctuates, these assets still benefit the wider European economy. Integrating energy considerations into the initial policy framework also ensures that the distributional impacts remain subject to public debate rather than being obscured. Energy is a necessary condition for data centre development, not a secondary concern. Addressing this early avoids a scenario where the implications of data centre buildout are presented as *fait accompli* after the fact, with costs falling disproportionately to the taxpayers. Moreover, before expanding the new data centre capacity with long-ranging implications, it would be useful to map out to what extent the current capacity is used efficiently, merely blocked or reserved for strategic purposes.³⁴

The public interest is not synonymous with an 'Nvidia arms race'. There is a danger of goal displacement, where the interests of private intermediaries in making profit and capturing emerging market opportunities are misaligned with the public interest in deploying technology sensibly and productively, with respect for societal boundaries. The success of efficiency-first models (such as DeepSeek) suggests that compute-constrained development can be more cost-effective and sustainable (Varoquaux, Luccioni & Whittaker 2024). Policymakers should be sceptical of industry claims that every new chip generation renders previous infrastructure obsolete and requires new investments. They should also be wary of becoming systematically entangled with corporate interests and providing a sovereign buffer to facilitate production decisions by private companies. A strategic choice would be **procuring chips once the depreciation curve sets in**. There will be

³² While we maintain concerns about AI concentration dynamics, this analysis focuses specifically on infrastructure policy implementation challenges.

³³ This has been discussed also in previous work. See Warso 2026, as well as Renda & Kyosovska 2025.

³⁴ Some recent research suggests that the data centre capacity in some member states might be operating "nowhere near the full capacity", with estimates that for the Netherlands, up to 2/3 of the available maximum capacity is reserved, but unused. See Schulze & van Veen 2025a.

a flood of chips in the market in the years to come. By waiting to build out data centres first and securing chips later, the EU could reduce the cost tenfold.

A strategic diversification of chip providers is essential to grant the public sector the agency needed to prevent vendor capture and maintain long-term technological autonomy. One source of agency in the AI value chain is **silicon diversity**, which strategically prevents capture and vendor-lock in by one provider (Boulton 2023). European companies have shown some appetite to develop alternatives, for instance the deployment of Groq chips in Equinix data centres in Helsinki, and Cerebras chips by European AI companies like Aleph Alpha and Cerebras (Dahlstroem & Stephens 2025).

Compute demand and supply must be pooled on a centralised platform to prevent market fragmentation and capture by the existing players.³⁵ European providers are currently fragmented, with many competing for little demand. This makes them vulnerable to being drawn into the pull of hyperscaler demand and then becoming commodified providers as part of private aggregators such as Nvidia Lepton DGX. Addressing the concentration of demand requires explicitly coordinating demand between users, competitive unbundling within cloud markets and controlled integration through interoperability.

Conclusion

This report highlights the market dynamics at play ahead of the EU Cloud and AI Development Act and the European efforts to stay in the AI race. Without strategic perspective, the default direction of initiatives like tripling the data centre capacity in the EU will entrench the incumbent major market players in a captured ecosystem. Crafting an alternative requires a decisive, strategic approach that charts a conscious path in the thicket of various private interests.

Although we have set aside the question of whether the AI buildout is justified in the first place, there is growing, tangible evidence that its current market trajectory may exacerbate inequality, constrain innovation, negatively impact the environment and weaken accountability in society. These are pressing matters. While the political choices are for elected politicians to make, this report highlights several ways in which policy decisions risk being influenced and constrained by the material realities of the AI buildout. We hope that acknowledging these constraints provides a path for a fact-based debate on the kinds of societies Europe wants to build with AI.

³⁵ For proposals in this vein, see Schulze & Van Veen 2025b.

References

- Arun, A. (2025): Bubble or Nothing. Data Center Project Finance, Center for Public Enterprise, November 2025. <https://publicenterprise.org/wp-content/uploads/Bubble-or-Nothing.pdf>.
- Berthelot, B. (2026): Fluidstack Drops Out of Marquee €10 Billion AI Project in France, in: Bloomberg, 18.03.2026. <https://www.bloomberg.com/news/articles/2026-03-18/fluidstack-drops-out-of-marquee-10-billion-ai-project-in-france>.
- Bobrowsky, M. (2025): Big Tech Is Spending More Than Ever on AI and It's Still Not Enough, in: The Wall Street Journal, 30.10.2025. <https://www.wsj.com/tech/ai/big-tech-is-spending-more-than-ever-on-ai-and-its-still-not-enough-f2398cfe>.
- Boonstra, A. (2023): The beginning of the end for FLAP?, in: DatacenterDynamics, 27.03.2023. <https://www.datacenterdynamics.com/en/opinions/the-beginning-of-the-end-for-flap/>.
- Boulton, C. (2023): How generative AI is accelerating silicon diversity, CIO, 8.08.2023. <https://www.cio.com/article/648988/how-generative-ai-is-accelerating-silicon-diversity.html>.
- Bradshaw, T. (2025): Google's 'TPU' chip puts OpenAI on alert and shakes Nvidia investors, in: Financial Times, 8.12.2025. <https://www.ft.com/content/d8585870-17a5-43a0-95ef-cbeb1995107>.
- Brennan, K., Kak, A. and Myers West, S. (2025): Artificial Power. 2025 Landscape Report, AI Now Institute, 3.06.2025. https://ainowinstitute.org/wp-content/uploads/2025/06/FINAL-20250609_AINowLandscapeReport_Full.pdf.
- Brookfield (2025a): Building the Backbone of AI, August 2025. https://www.brookfield.com/sites/default/files/documents/Brookfield_Building_the_Backbone_of_AI.pdf.
- Brookfield (2025b): Brookfield Launches \$100 Billion AI Infrastructure Program, 19.12.2025. <https://bam.brookfield.com/press-releases/brookfield-launches-100-billion-ai-infrastructure-program>.
- Business Wire (2025): Fluidstack Delivering Europe's Largest AI Supercomputer to Mistral AI in 2025, 3.03.2025. <https://www.businesswire.com/news/home/20250303970989/en/Fluidstack-Delivering-Europes-Largest-AI-Supercomputer-to-Mistral-AI-in-2025>.
- Butler, G. (2024): Microsoft to invest \$7.16bn in new data centers in Aragon, Spain, in: DatacenterDynamics, 18.06.2024. <https://www.datacenterdynamics.com/en/news/microsoft-to-invest-716bn-in-new-data-centers-in-aragon-spain/>.
- Butler, G. (2025): Chipping away at the economics of neoclouds, in: DatacenterDynamics, 9.10.2025. <https://www.datacenterdynamics.com/en/analysis/chipping-away-at-the-economics-of-neoclouds/>.
- Butler, G. (2026): Brookfield challenges cloud giants with new offering, in: DatacenterDynamics, 5.01.2026. <https://www.datacenterdynamics.com/en/news/brookfield-challenges-cloud-giants-with-new-offering/>.
- CBRE (2023): The impact of the AI Boom on the European Data Centre Market, October 2023. <https://mktgdocs.cbre.com/2299/20720f67-81c1-4d1c-b262-918199ea2d-bc-539424452.pdf>.
- CBRE (2025a): European Real Estate Market Outlook 2025: Chapter 8, 13.01.2025. <https://www.cbre.com/insights/books/european-real-estate-market-outlook-2025>.
- CBRE (2025b): European Data Centres, figures Q3 2025. https://mktgdocs.cbre.com/2299/356288aa-0100-4b2c-b519-40c0228d5278-170902235/European_Data_Centres_Figures_.pdf.
- CBRE (2025c): Business Insights. The Impact of AI on Europe's Data Centre Market, 13.10.2025. <https://www.cbre.co.uk/insights/articles/the-impact-of-ai-on-europes-data-centre-market>.
- CBRE (2025d): European Real Estate Market Outlook 2025, n.d. <https://mediaassets.cbre.com/-/media/project/cbre/shared-site/insights/books/european-real-estate-market-outlook-2025/european-real-estate-market-outlook-2025.pdf?rev=62ea6b-b8eeef48a4814dc789ab8c8f73>.
- Chen, K., Patel, D., Nishball, D. and Chiam, I. (2025): AMD vs NVIDIA Inference Benchmark: Who Wins? – Performance & Cost Per Million Tokens, in: SemiAnalysis, 23.05.2025. <https://newsletter.semianalysis.com/p/amd-vs-nvidia-inference-benchmark-who-wins-performance-cost-per-million-tokens>.
- Clark, A. (2025): Amazon Completes AI Supercomputer with Anthropic. What It Means for the Stock, Barron's, 29.10.2025. <https://www.barrons.com/articles/amazon-stock-price-anthropic-ai-7b49d063>.
- Cooper, R. (2025): How advanced technologies are helping AI Data Centres keep their cool, CBRE, 19.03.2025. <https://www.cbre.co.uk/insights/articles/how-advanced-technologies-are-helping-ai-data-centers-keep-their-cool>.
- Craske, B. (2025): How Will OpenAI Build AI Gigafactory Data Centre in Norway?, in: DataCentre Magazine, 1.8.2025. <https://datacentremagazine.com/news/how-will-openai-build-ai-gigafactory-data-centre-in-norway>.
- Dahlstrom, R. and Stephens, C. (2025): Groq and Equinix Partner for Low-Latency AI Infrastructure in Europe, Equinix, 20.08.2025. <https://blog.equinix.com/blog/2025/08/20/groq-and-equinix-partner-for-low-latency-ai-infrastructure-in-europe/>.
- Dave, P. (2024): How a Scrappy Cryptominer Transformed Into the Multibillion-Dollar Backbone of the AI Boom, in: Wired Magazine, 3.05.2024. <https://www.wired.com/story/coreweave-scrappy-cryptominer-multibillion-dollar-ai/>.
- Deutsche Telekom (2025): AI Turbo for Gigafactories: Telekom announces European Industrial AI Infrastructure with NVIDIA, 13.06.2025. <https://www.telekom.com/en/media/media-information/archive/ai-turbo-nvidia-and-deutsche-telekom-1093532>.
- Donegan, M. (2025): Telefónica leads Spain's bid for EU-backed AI Gigafactory, TelcoTitans, 25.06.2025. <https://www.telcotitans.com/telefonicawatch/telefonica-leads-spains-bid-for-eu-backed-ai-gigafactory/9385.article>.
- Donnelly, C. (2025): CBRE charts rise of neocloud providers within European colocation market, in: Computer Weekly, 13.11.2025. <https://www.computerweekly.com/news/366634475/CBRE-charts-rise-of-neocloud-providers-within-European-colocation-market>.
- Efrati, A. (2025): Nvidia In Talks to Lease Chips to OpenAI as Startup's Server Spending Plan Rises to \$1 Trillion, The Information, 23.9.2025. <https://www.theinformation.com/briefings/nvidia-talks-lease-chips-openai-startups-server-spending-plan-rises-1-trillion>.
- EUDCA (2025): State of European Data Centres 2025. Revision 1.0, May 2025. <https://www.eudca.org/state-of-european-data-centres-2025>.
- European Commission (2026): Shaping Europe's leadership in artificial intelligence with the AI continent action plan, n.d. https://commission.europa.eu/topics/competitiveness/ai-continent_en.
- Gabert-Doyon, J. (2026): There might be good reasons for Britain to build data centres. Job creation isn't one of them, in: Financial Times, 28.1.2026. <https://www.ft.com/content/1a48f0db-b667-4746-83eb-0c4c04e2584f>.
- Gardizy, A. and Efrati, A. (2025): Sam Altman Wants 250 Gigawatts of Power. Is that Possible?, in: The Information, 29.09.2025. <https://www.theinformation.com/newsletters/ai-infrastructure/sam-altman-wants-250-gigawatts-power-possible>.
- Gerlat, P. (2025): Mistral AI and Nvidia: Towards a Sovereign AI Infrastructure in Europe, Actulia, 13.06.2025. <https://www.actulia.com/en/news/mistral-ai-and-nvidia-towards-a-sovereign-ai-infrastructure-in-europe/>.
- Gröger, J., Behrens, F., Gailhofer, P. and Hilbert, I. (2025): Umweltauswirkungen Künstlicher Intelligenz, Öko-Institut Berlin for Greenpeace e.v. Hamburg, 12.05.2025. <https://www.oeko.de/publikation/umweltauswirkungen-kuenstlicher-intelligenz/>.
- Groq (2025): Groq and Nvidia Enter Non-Exclusive Inference Technology Licensing Agreement to Accelerate AI Inference at Global Scale, 24.12.2025. <https://groq.com/newsroom/groq-and-nvidia-enter-non-exclusive-inference-technology-licensing-agreement-to-accelerate-ai-inference-at-global-scale>.
- Hawkins, Z., Lehdonvirta, V. and Wu, B. (2026): AI Compute Sovereignty: Infrastructure Control Across Territories, Cloud Providers, and Accelerators, SSRN, 17.04.2026. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5312977.

- Hess, J. and Sieker, F. (2025): Built for Purpose? Demand-Led Scenarios for Europe's AI Gigafactories, Interface and Bertelsmann Stiftung, 22.10.2025. <https://www.interface-eu.org/publications/ai-gigafactories>.
- Hu, T. (2016): A Prehistory of the Cloud, The MIT Press.
- Huang, R., Clark, K. and Jin, B. (2026): Google Is Exploring Ways to Use Its Financial Might to Take On Nvidia, in: The Wall Street Journal, 20.2.2026. <https://www.wsj.com/tech/ai/google-is-exploring-ways-to-use-its-financial-might-to-take-on-nvidia-0fbadc84>.
- Iacurci, G. (2025): AI data center 'frenzy' is pushing up your electric bill – here's why, CNBC, 26.11.2025. <https://www.cnbc.com/2025/11/26/ai-data-center-frenzy-is-pushing-up-your-electric-bill-heres-why.html>.
- Iyer, R. (2025): Black Forest Labs raises \$300M at \$3.25B valuation, in: TechCrunch, 1.12.2025. <https://techcrunch.com/2025/12/01/black-forest-labs-raises-300m-at-3-25b-valuation/>.
- James, L. (2025): Data centers turn to commercial aircraft jet engines bolted onto trailers as AI power crunch bites – cast-off turbines generate up to 48 MW of electricity apiece, Tom's Hardware, 22.10.2025. <https://www.tomshardware.com/tech-industry/data-centers-turn-to-ex-airliner-engines-as-ai-power-crunch-bites>.
- Kaesebier, M., Rugamer, M. and Marchandon, L. (2025): Deutsche Telekom, Nvidia to launch \$1.2 billion AI cloud for industrial clients, Reuters, 4.11.2025. <https://www.reuters.com/business/media-telecom/deutsche-telekom-partners-with-nvidia-ai-cloud-q1-2026-2025-11-04/>.
- Kak, A. and Myers West, S. (eds.) (2024): AI Nationalism(s): Global Industrial Policy. Approaches to AI, AI Now Institute, March 2024. <https://ainowinstitute.org/wp-content/uploads/2024/03/AI-Nationalisms-Global-Industrial-Policy-Approaches-to-AI-March-2024.pdf>.
- Kedrosky, P. (2024): What If AI Tokens Are Too Cheap to Meter?, 22.9.2024. <https://paulkedrosky.com/what-if-ai-tokens-are-too-cheap-to-meter/>.
- Lehdonvirta, V., Wu, B., and Hawkins, Z. (2024): Compute North vs. Compute South: The Uneven Possibilities of Compute-based AI Governance Around the Globe, in: Proceedings of the AAAI/ACM Conference on AI, Vol. 7, pp. 828-838, 16.10.2024. <https://ojs.aaai.org/index.php/AIES/article/view/31683>.
- Lehdonvirta, V., WU, B., and Hawkins, Z. (2023): Weaponized interdependence in a bipolar world: How economic forces and security interests shape the global reach of U.S. and Chinese cloud data centres, SSRN, 20.12.2023. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4670764.
- Leppert, R. (2025): What we know about energy use at U.S. data centers amid the AI boom, Pew Research Center, 24.10.2025. <https://www.pewresearch.org/short-reads/2025/10/24/what-we-know-about-energy-use-at-us-data-centers-amid-the-ai-boom/#how-could-data-centers-affect-americans-electricity-bills>.
- Lopatto, E. (2023): For this startup, Nvidia GPUs are currency', in: The Verge, 8.8.2023. <https://www.theverge.com/2023/8/8/23824661/coreweave-nvidia-debt-gpu-ai-chips-collateral>.
- Lopatto, E. (2025): I looked into CoreWeave and the abyss gazed back, in: The Verge, 17.11.2025. <https://www.theverge.com/ai-artificial-intelligence/822011/coreweave-debt-data-center-ai>.
- Martindale, J. (2025): GPU depreciation could be the next big crisis coming for AI hyperscalers – after spending billions on buildouts, next-gen upgrades may amplify cash-flow quirks, Toms Hardware, 18.11.2025. <https://www.tomshardware.com/tech-industry/gpu-depreciation-could-be-the-next-big-crisis-coming-for-ai-hyperscalers-after-spending-billions-on-buildouts-next-gen-upgrades-may-amplify-cashflow-quirks>.
- Mazza, M., Sachdeva, P., Arutyunyan, S. and Alatovic, T. (2025): The evolution of neo-clouds and their next moves, McKinsey & Company, 19.11.2025. <https://www.mckinsey.com/capabilities/tech-and-ai/our-insights/the-evolution-of-neoclouds-and-their-next-moves#/>.
- McKenna, B. (2025): Nvidia's "Aqui-Hire" of Groq Eliminates a Potential Competitor and Marks Its Entrance Into the Non-GPU, AI Inference Chip Space, in: The Motley Fool, 28.12.2025. <https://www.fool.com/investing/2025/12/28/nvidia-groq-deal-acquisition-ai-inference-lpu/>.
- Mukherjee, S. and Marchandon, L. (2026): France's Mistral raises \$830 million in debt for AI data centre build-up, Reuters, 30.3.2026. <https://www.reuters.com/business/finance/frances-mistral-raises-830-million-debt-ai-data-centre-build-up-2026-03-30/>.
- Nokia (2025): Nokia partners with Nvidia, 28.10.2025. <https://www.nokia.com/newsroom/nokia-partners-with-nvidia/>.
- Nvidia (2025): NVIDIA DGX Cloud Lepton Connects Europe's Developers to Global NVIDIA Compute Ecosystem, 11.06.2025. <https://nvidianews.nvidia.com/news/nvidia-dgx-cloud-lepton-connects-europes-developers-to-global-nvidia-compute-ecosystem>.
- Nvidia (2026): Siemens and NVIDIA Expand Partnership to Build the Industrial AI Operating System, 6.1.2026. <https://nvidianews.nvidia.com/news/siemens-and-nvidia-expand-partnership-industrial-ai-operating-system>.
- OECD (2023): A blueprint for building national compute capacity for artificial intelligence, OECD Digital Economy Papers, 28.2.2023. Doi: <https://doi.org/10.1787/876367e3-en>.
- OECD (2025a): Measuring domestic public cloud compute availability for artificial intelligence, OECD Artificial Intelligence Papers No. 49, October 2024. https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/10/measuring-domestic-public-cloud-compute-availability-for-artificial-intelligence_39fa6b0e/8602a322-en.pdf.
- OECD (2025b): Competition in the Provision of Cloud Computing Services – Note by the European Union (DAF/COMP/WD(2025)39), 20.05.2025. [https://one.oecd.org/document/DAF/COMP/WD\(2025\)39/en.pdf](https://one.oecd.org/document/DAF/COMP/WD(2025)39/en.pdf).
- Patel, D. and Nishball, D. (2024): AI Neocloud Playbook and Anatomy, in: SemiAnalysis, 3.10.2024. <https://newsletter.semianalysis.com/p/ai-neocloud-playbook-and-anatomy>.
- Patel, D., Nishball, D. and Eliahou Ontiveros, J. (2024): AI Datacenter Energy Dilemma – Race for AI Datacenter Space, in: SemiAnalysis, 13.3.2024. <https://newsletter.semianalysis.com/p/ai-datacenter-energy-dilemma-race>.
- Renda, A. and Kyosovska, N. (2025): EU plans for AI (giga)factories: sanctuaries of innovation, or cathedrals in the desert?, CEPS, 3.11.2025. <https://www.ceps.eu/ceps-publications/eu-plans-for-ai-gigafactories-sanctuaries-of-innovation-or-cathedrals-in-the-desert/>.
- Reuters (2025): CoreWeave, Nvidia sign \$6.3 billion cloud computing capacity order, 15.09.2025. <https://www.reuters.com/business/coreweave-nvidia-sign-63-billion-cloud-computing-capacity-order-2025-09-15/>.
- Rogers, O. (2025): Neoclouds: a cost-effective AI infrastructure alternative, Uptime Institute, 26.2.2025. <https://journal.uptimeinstitute.com/neoclouds-a-cost-effective-ai-infrastructure-alternative/>.
- Saari, L. (2024): Competition Policy and Large-Scale AI Markets in Europe, pp. 44-51, in: Oosterwijk, G., Hasdenteufel, M. and Nogarede, J. (eds.): Time to build a European digital ecosystem. Recommendations for the EU's digital policy, FEPS-FES Policy Study, December 2024. <https://feps-europe.eu/publication/time-to-build-a-european-digital-ecosystem-2/>.
- Schulze, M. and Van Veen, C. (2025a): Four Key Conclusions about Data Center Environmental Transparency in the Netherlands, Leitmotiv, 21.11.2025. <https://leitmotiv.digital/publications/four-key-conclusions-data-center-environmental-transparency-neth>.
- Schulze, M. and Van Veen, C. (2025b): Why Europe Doesn't Need New Infrastructure – Just an Open Cloud Market, Leitmotiv, 12.12.2025. <https://leitmotiv.digital/publications/why-europe-doesnt-need-new-infrastructure-just-open-cloud-market>.
- SemiAnalysis (2025) [@SemiAnalysis_]: BREAKING: A couple of months ago, Neo-Clouds were scared that DGX Lepton was going to commoditize their margins. This is no longer the case, X/Twitter, 31.10.2025, 1:14. https://x.com/SemiAnalysis_/status/1984051218688204827.
- Serrano, A. (2025): Banco de Fomento proposes construction of mega AI data center in Sines with an investment of 4 billion euros, Alentrium, 21.06.2025. <https://alentrium.pt/en/banco-de-fomento-propoe-construcao-de-mega-data-center-de-ia-em-sines-com-investimento-de-4-mil-milhoes-de-euros/>.
- Shrivastava, A., Salazar, L., Gaffey, B., Grundin, G., Cubela, S., Lajous, T., Frade, M., Agarwal, A., Orimoloye, D. and Shaw, N. (2025): AI infrastructure: A new growth avenue for telco operators, 28.2.2025. <https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/ai-infrastructure-a-new-growth-avenue-for-telco-operators#/>.
- Siemens (2026): Siemens and NVIDIA Expand Partnership to Build the Industrial AI Operating System, 6.1.2026. <https://press.siemens.com/global/en/pressrelease/siemens-and-nvidia-expand-partnership-build-industrial-ai-operating-system>.
- Smyth, O. (2025): Government warned of rising household bills as data centres strain grid, RTE, 21.8.2025. <https://www.rte.ie/news/primetime/2025/0821/1529518-govt-warned-of-rising-household-bills-as-data-centres-strain-grid/>.
- Swinhoe, D. (2025): MGX, Bpifrance, Nvidia, and Mistral AI plan 1.4GW Paris data center campus, in: DatacenterDynamics, 20.5.2025. <https://www.datacenterdynamics.com/en/news/mgx-bpifrance-nvidia-and-mistral-ai-plan-14gw-paris-data-center-campus/>.
- Tairych, D. and Delp, A. (2025): What is a Neocloud?, Equinix, 14.10.2025. <https://blog.equinix.com/blog/2025/10/14/what-is-a-neocloud/>.
- Thole, H. (2025): 'AI-zoekmachine voor materialen' CuspAI haalt 100 miljoen op – Volkswagen vergoedt kopers van 100.000 sjoemeldiesels, MT/Sprout, 11.9.2025. <https://mtsprout.nl/nieuws/cuspai-100-miljoen-dollar>.
- Varoquaux, G., Luccioni, A. and Whittaker, M. (2024): Hype, Sustainability, and the Price of the Bigger-is-Better Paradigm in AI, arxiv, 21.9.2024. <https://doi.org/10.48550/arXiv.2409.14160>.

Verbovnikov, N. (2026): Top cloud providers in Europe, Innwise, 12.2.2026. <https://innwise.com/blog/top-cloud-providers-in-europe/>.

Walton, J. (2025): Nvidia shows off Rubin Ultra with 600,000-Watt Kyber racks and infrastructure, coming in 2027, Tom's Hardware, 19.3.2025. <https://www.tomshardware.com/pc-components/gpus/nvidia-shows-off-rubin-ultra-with-600-000-watt-kyber-racks-and-infrastructure-coming-in-2027>.

Warso, Z. (2026): Who Controls Europe's AI Future? How the Gigafactories Bet Shapes the EU's AI Trajectory, Open Future Foundation, 9.2.2026. <https://openfuture.eu/blog/who-controls-europes-ai-future/>.

Zitron, E. (2025): The Case Against Generative AI, in: Where's Your Ed At?, 29.9.2025. <https://www.wheresyoured.at/the-case-against-generative-ai/#there-is-less-than-a-billion-dollars-in-ai-compute-revenue-outside-of-nvidia-hyperscalers-and-openai-and-nvidia-is-using-its-compute-deals-to-help-neoclouds-raise-more-debt-to-buy-more-gpus>.

Zografos, J. (2025): Data Center Alley: How Ashburn became a data center hub of the world, Digital Realty, 7.1.2025. <https://www.digitalrealty.com/resources/blog/northern-virginia-ashburn-data-centers>.

About the authors



Leevi Saari is a Doctoral Researcher at the University of Amsterdam and an EU Policy Fellow at the AI Now Institute. In these roles, he examines the political economy of artificial intelligence within the European Union. A frequent public commentator on European AI developments, Leevi has recently published on the capital ecosystems behind European AI startups, competition dynamics in the model layer, and corporate strategies navigating geostrategic tensions. Previously, he served as a Policy Advisor in the European Parliament, where he worked on data, AI, and cloud legislation.



Boxi Wu is a Doctoral Researcher at the University of Oxford where their research investigates the social and political impacts of AI infrastructure. They are currently a Research Fellow at the AI Now Institute and a member of the Aalto-Oxford University Digital Economic Security Lab. Boxi regularly contributes to AI policy research and has published widely on the ethics and governance of AI and digital infrastructure. Boxi previously worked in AI ethics at Google DeepMind.

Acknowledgements

We thank our reviewers for their useful comments and suggestions for improvement. The public reviewers are:

Advait Arun, Robin Berjon, Frederike Kaltheuner, Vili Lehdonvirta, Kaarlo Liukkonen, Julia Carver, Justin Nogarede, Lloyd Bingham, Max Schulze, Elisabeth Siegel, Stefkla Schmid, Max von Thun, Zuzanna Warso

Chasing the AI Cloud in Europe

This publication analyses the emergence of alternative AI computing providers – so-called “neoclouds” – that are heralded as alternatives to US hyperscalers. Who are they? What are their market strategies, and to what extent can they carry EU policy-makers’ aspirations for greater technological autonomy in AI? This paper provides answers.

With a detailed analysis of the financial fundamentals, market dynamics, and energy constraints shaping AI cloud compute in Europe, and neoclouds in particular, the authors show that EU ambitions for “technological sovereignty in AI” are fraught with hard trade-offs that deserve public debate, as well as risks that require an industrial policy approach more specific than just “tripling data centre capacity in the EU”. In its absence, EU policy on AI risks creating a massive public backlash and may well end up subsidising US AI cloud providers while further deepening the EU’s dependence.

The paper ends with several policy recommendations, most notably to prioritise energy and grid investments, procure chips behind the depreciation curve, diversify chip providers, and finally pool compute demand and supply on a centralised platform.

Further information on the topic can be found here:

➔ [fes.de](https://www.fes.de)