

Technology Report

Sovereign AI Is the Next Fault Line in the Global Tech Sector

The electronics supply chain was only the start of tech's global decoupling.

By Anne Hoecker, Jonathan Frick, Jue Wang, Balaji Thirumalai, and Karen Harris

6 min read



Listen to this article 8 minutes

At a Glance

- ▶ Sovereign AI blocs are emerging as governments worldwide spend billions of dollars subsidizing domestic computing infrastructure and AI models.
- ▶ Locally based data center providers account for nearly a quarter of new computing capacity coming online in the next few years.
- ▶ Despite sovereign AI momentum, tech incumbents' global scale and deep coffers provide significant advantages over domestic competitors.
- ▶ Data center operators will enjoy a short-term windfall, but there's a real risk of overcapacity.

This article is part of Bain's 2024 Technology Report.

[EXPLORE THE REPORT](#)

As technology companies race to capitalize on breakthroughs in large language models (LLMs) and generative artificial intelligence (AI), executives must now grapple with an additional layer of complexity and opportunity: the emergence of “sovereign” AI blocs around the world.

De-globalization in technology began with the electronics supply chain, particularly semiconductors. Disruptions from Covid-19 and geopolitical tensions between the US and China (including export controls and restrictive policies on trade and talent) pushed tech companies to rapidly invest in making their supply chains more resilient. They've expanded their manufacturing footprints beyond China and created more flexibility within their talent pools. With government support, companies are building new semiconductor hubs in places including the US, India, Germany, and Japan.

Now the post-globalization movement in technology is spreading to data, AI, security, and privacy. Governments worldwide—including India, Japan, France, Canada, and the United Arab Emirates—are spending billions of dollars to subsidize sovereign AI. In other words, they're investing in domestic computing infrastructure and AI models developed within their borders, trained on local data and languages.

While it's tempting to compare sovereign AI to the decoupling of semiconductor supply chains, the challenges are quite different. For example, compared to the semiconductor market, which has a complex supply chain with intellectual property fragmented throughout, the AI market is easier to enter. This is largely due to open-source LLMs, which make launching new AI products simpler.

As the sovereign AI push picks up steam, several factors will determine how it plays out.

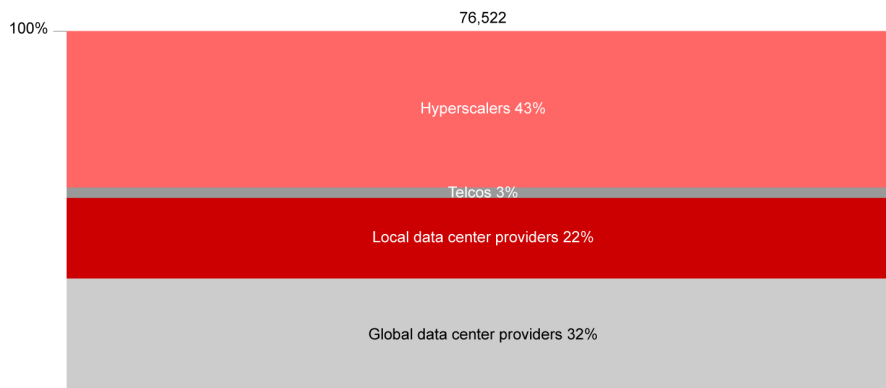
Factors favoring sovereign AI

1. **National interests:** Governments view localized AI as critical for protecting data privacy, ensuring national security, building or strengthening domestic high-tech ecosystems, and growing their economies. Countries can't afford to fully rely on others for AI and cloud computing capabilities due to the economic value at stake and the decoupling of the countries leading the AI race—the US and China.
2. **Infrastructure:** Like any utility, physical infrastructure for AI and cloud computing must be built somewhere and will require massive capital investments in data centers, computing capacity, and electrical grids. These investments intersect with other national infrastructure issues like the green transition in the electricity grid, which AI's significant power demand will complicate. Locally based data center providers account for nearly a quarter of new computing capacity coming online in the next few years, while technology hyperscalers are planning to add the most (see Figure 1). It's also notable that national governments have ordered at least 40,000 graphics processing units (GPUs) themselves over the past year.

Figure 1

Nearly a quarter of new data center capacity will come from local providers, with hyperscalers planning the most capacity

Global forecasted new data center capacity through 2027, by provider (MW)



Note: Global providers operate on more than two continents
Sources: IDC 2023 Datacenter Deployment and Spend Forecast, 1H 2023; Bain analysis

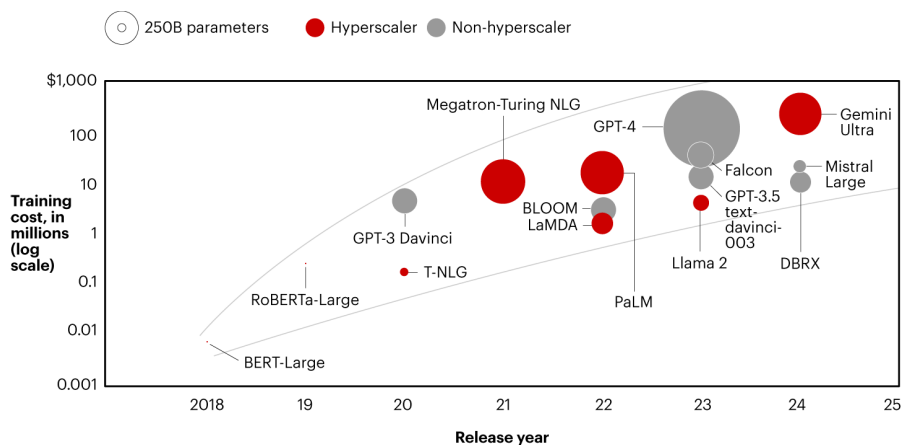
- 3. Regulatory strategies:** AI regulatory strategies are diverging across borders. The leading AI markets—the US, EU, and China—are taking very different approaches so far.
- 4. Localization:** Many AI models will need to be specific to local languages and context. Some applications will differ across countries to comply with security and privacy regulations and meet local market needs. AI use cases in healthcare, education, and agriculture, for example, will vary greatly between developed and emerging economies.

Factors working against sovereign AI

- 1. Scope of subsidies:** Thus far, governments haven't subsidized national AI initiatives to the extent seen with semiconductor fabs or to the degree likely required to nurture local champions that could compete at scale with global incumbents. (The powerful open-source LLM series Falcon, backed by hundreds of millions of dollars from an arm of the United Arab Emirates government, is a notable exception.)
- 2. Global scale:** This still provides critical advantages for developing a winning AI platform, including network effects (e.g., access to a large developer ecosystem), deep coffers, and the ability to spread R&D costs across worldwide operations. LLM training costs have grown exponentially over the past few years, with the most expensive models exceeding \$100 million (see Figure 2). Although smaller, more cost-efficient models are also being released, the cost dynamics continue to favor large global firms.

Figure 2

Large language model training costs have increased exponentially, but smaller, more cost-efficient models are also being released



Notes: Training cost estimates exclude staff costs; Mistral Large parameters are assumed to be equal to Mistral 8x7B due to similar capabilities; GPT-3.5 text-davinci-003 parameters are assumed to be same as GPT-3.5
Sources: Epochai.org; news articles; analyst reports; company websites; research papers; Bain analysis

- Incumbents' adaptation:** Global tech companies are adapting to governments' push for sovereign AI by localizing operations, complying with local rules, and forming joint ventures with local firms.
- Practical realities:** Aspiring domestic competitors must navigate the same practical realities as multinational companies: significant investments in securing land, regulatory approvals, power, connectivity, and other key elements for AI initiatives.

Takeaways for executives

Establishing successful sovereign AI ecosystems will be time-consuming and incredibly expensive. While less complex in some important ways than building semiconductor fabs, these projects require more than securing local subsidies.

Hyperscalers and other big tech firms may continue to invest in localized operations. This could fragment their ecosystems and R&D globally, though their scale will remain a significant advantage.

New AI workloads and fragmentation created by sovereignty could enable AI challengers to reach hyperscale. These challengers will need to recognize the power of the current hyperscaler ecosystem and prioritize business opportunities that capitalize on their competitive advantages, while partnering with big tech companies where possible.

Data center operators and hardware suppliers will enjoy a short-term windfall as companies and governments splurge on computing capacity. Nvidia, for example, projected \$10 billion in revenue from governments' sovereign AI investments in 2024, up from zero last year. However, data center owners risk overcapacity, similar to telecom networks in the early 2000s. Suppliers of silicon and other hardware may see accelerated growth rates level off long-term.

Lastly, investors have a chance to stake high-value claims in a hot asset class, including new sub-asset classes. For example, secured financing tied to GPUs is becoming a more common form of corporate debt. Successful investors

will base bets on a well-defined risk/return profile, deciding between lower-risk investments in “picks and shovels” like GPUs and data centers or higher-risk/higher-reward investments such as LLMs and cloud platforms.

[Read the Next Chapter](#)

Five Functions Where AI Is Already Delivering



Read our 2024 Technology Report

[DOWNLOAD THE PDF](#)

[EXPLORE THE REPORT](#)

More from the report

Commercializing Innovation	▶
Investing in a Shifting Market	▶
Scope M&A	▶
Sovereign AI	▶
Top AI Functions	▶
The AI Opportunity	▶
Big and Small Computing	▶
The AI Chip Shortage	▶
The Software Cycle	▶
AI in Tech Services	▶
Deploying Generative AI	▶
Software Development Efficiency	▶
Customer Success	▶
AI Everywhere	▶

Published in September 2024

VectorSM is a service mark of Bain & Company, Inc.