Natural Language Processing (2025), **31**, pp. 1312–1321
doi:10.1017/nlp.2025.10007

CAMBRIDGE UNIVERSITY PRESS

INDUSTRY WATCH

Sovereign AI in 2025

Robert Dale

Language Technology Group, Sydney, Australia
Email: rdale@language-technology.com

(Received 14 July 2025; revised 14 July 2025; accepted 15 July 2025; first published online 4 August 2025)

Abstract

At the London Tech Week event in early June, Nvidia CEO Jensen Huang praised the UK as the 'envy of the world' when it comes to AI researchers, but he also criticised it as the largest AI ecosystem in the world without its own infrastructure. The criticism is somewhat self-serving: when the UK does get around to building out that infrastructure, it's certain to consist largely of chips sold by Huang's company. It's also unsurprising: Huang has been pitching the idea of 'sovereign AI' since at least 2023, conscious that nation states are the next deep pockets to target after the hyperscalers and generously funded model builders. In a world where the only real contenders in the race for AI supremacy are the US and China, we look at how the pursuit of AI sovereignty is playing out across the rest of the planet.

Keywords: sovereignty; AI

1. Introduction

The concept of sovereign AI – the idea that nations should develop and maintain independent AI capabilities rather than rely on systems controlled by foreign entities or multinational corporations – started gaining prominence around 2020, particularly within European policy discussions focused on digital sovereignty and data protection. But it was Nvidia CEO Jensen Huang who elevated the visibility of the concept, starting with an Nvidia earnings call in November 2023. There's been a rush of activity across the globe in the name of AI sovereignty since then.

The general notion of sovereignty refers to a nation state's ability to govern itself without external interference. Applied to AI, sovereignty entails the ability to develop AI technologies on domestically based infrastructure, using local data and human talent, free from dependency on foreign platforms or corporations. True sovereignty implies control over the full AI lifecycle, from the collection of data and training of models through to the deployment and governance of AI systems.

There are a number of motivations for a country to seek AI sovereignty, including at least the following:

- **national security**: reliance on systems developed overseas can result in vulnerabilities in critical infrastructure and may result in exposure to backdoors or influence operations;
- **data privacy and control**: keeping data at home protects against foreign laws and data misuse;
- **economic competitiveness**: developing domestic capabilities fosters a local AI ecosystem and helps avoid technological lock-in;

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/nlp.2025.10007 Published online by Cambridge University Press

- **regulatory autonomy**: if global AI standards are set by dominant foreign players, local priorities around transparency, accountability and fairness may be sidelined;
- **cultural and linguistic representation**: AI trained on local data provides better support for local knowledge, education and services;
- **strategic autonomy**: in times of geopolitical conflict or trade wars, it's dangerous to rely on external suppliers for essential functions.

The relative importance of each of these drivers at any given time will vary across nations, given their particular economic and political contexts. So, how have countries in the various regions of the world responded to the challenge of AI sovereignty – and is AI sovereignty even achievable?

2. Full-stack sovereignty

As we defined the concept above, possessing AI sovereignty would appear to require complete self-sufficiency, at least from a technological perspective. There are a number of layers to the AI tech stack, so we might refer to this degree of sovereignty as 'full-stack' sovereignty. Although there are many components that go to make up a complete AI ecosystem, several in particular stand out as being particularly central:

- the data required for building AI models;
- the models themselves;
- the GPU clusters required for training and inference; and
- the cloud data centres that host the GPUs.

Each of these elements has its own supply chain and dependencies: data centres occupy physical real estate and require significant amounts of energy; chips have to be designed, and their subsequent fabrication requires sophisticated photolithography machines; model development and deployment depends on research and the researchers who carry out that research; and human talent in turn depends on up-to-date training and education. It's a complex web of dependencies.

Self-sufficiency in all of these areas is a tough call, even for the US and China, the two acknowledged AI superpowers engaged in a battle for global supremacy. Nvidia (US) and Huawei (China) dominate in chip design but outsource fabrication to third-party foundries, Taiwan Semiconductor Manufacturing Company (TSMC) and China's Semiconductor Manufacturing International Corporation (SMIC), respectively. The fabs in turn rely on the photolithography machines used to produce integrated circuits, those machines being made by ASML (the Netherlands) and China's Shanghai Micro Electronics Equipment (SMEE).

China plays host to native activity in every layer of the stack. The US, on the other hand, has no domestic manufacturers of complete lithography systems, although it is home to the makers of essential components in those systems, and indeed ASML itself relies on core technologies and subsystems made in the US. ASML, and therefore the US, is technologically ahead here: at the time of writing, ASML is the only provider of the Extreme Ultraviolet lithography (EUV) required for producing cutting-edge chips at 7 nm and below. This advantage percolates up the supply chain. Under pressure from the US, ASML is forbidden to sell its machines to China, giving TSMC and therefore Nvidia the lead in their respective domains.

On a strict reading of the concept, China is arguably the only truly AI-sovereign nation. But the US's ability to exercise influence over its external dependencies means that it also has de facto sovereignty. The ASML advantage means that the US is technically in the lead for the time being; Trump's tech czar David Sacks puts China 18–24 months behind. But China is gaining ground fast, and finding other workarounds for the handicaps put in place by the US.

https://doi.org/10.1017/nlp.2025.10007 Published online by Cambridge University Press

3. Infrastructure dependence

The flip side of sovereignty is dependence.

The most obvious layer of the AI stack where countries inevitably have to cede sovereignty is that of the chips used for model training and inference.

The chip space is effectively owned by US and Chinese players. Nvidia dominates AI GPU demand with around 80% global market share (estimates vary a bit depending on how you count), and while Huawei's share of the global market is small, it leads in China, with one source suggesting it owns 75% of the domestic market. Together with other US and Chinese companies, Nvidia and Huawei fulfill the vast bulk of the worldwide demand for chips. Although there are some significant smaller players in third countries – Tenstorrent in Canada, Samsung in South Korea and Graphcore in the UK stand out, along with a few others – they represent risky bets for such fundamental national infrastructure. Chip sovereignty looks to be beyond reach if you're not already an AI superpower.

What about the data centres that house the chips? Again the market is dominated by a number of US corporations, generally collectively referred to as the hyperscalers: Amazon Web Services (32% global market share in 2024), Microsoft Azure (23% market share) and Google Cloud (10% market share), plus a few other players – Alibaba Cloud and Tencent Cloud have a noticeable presence in the Asia-Pacific. Every few weeks there seems to be yet another announcement of a multi-billion-dollar data centre build in another country by one of the hyperscalers, invariably claiming to support the country's ambitions for sovereignty. The list in Table 1, based on a review of news coverage over the last two years, gives a flavour of the US hyperscalers' reach, but is almost certainly incomplete.

4. Sovereign responses

According to the OECD AI Policy Observatory, over 70 countries and territories have published official national AI strategies or plans. Below, we look region-by-region at efforts towards AI sovereignty made across the world over the last couple of years. Not every country is covered here, but the most significant blocs get some scrutiny.

We'll focus particularly on three areas: data centre builds, chips and AI model development. As already noted, there are many other elements in an AI ecosystem, but these are arguably the most important in terms of achieving the goals of AI sovereignty.

4.1 The European Union

AI sovereignty, and technical sovereignty more generally, has been a much talked-about topic in European policy circles in recent years, and concerns have been accelerated by the fear that Trump, as part of his confrontational stance towards Europe, might decide to weaponise critical technologies. In December 2024, the European Commission appointed Finland's Henna Virkkunen as the first-ever commissioner for Tech Sovereignty, Security and Democracy. European tech companies and industry groups have called on EU leaders to take radical action supporting homegrown digital infrastructure, reducing foreign tech dependence through procurement requirements and strategic funding. But there's also a realism around the degree to which existing dependencies can just be switched off and the need to integrate into global tech ecosystems where appropriate.

4.1.1 Data centres

In December 2024, the EU announced it would invest €750M to establish and maintain seven 'AI factories' across Europe, with matching member states' funding for a total of €1.5B, with the aim

https://doi.org/10.1017/nlp.2025.10007 Published online by Cambridge University Press

Table 1. Recent US hyperscaler investment outside the US, in US dollars – not an exhaustive list

Hyperscaler	Country	Amount (USD)	Announced
AWS	Australia	$13.3B	June 2025
AWS	Germany	$9.1B	May 2024
AWS	Italy	$1.3B	Early 2024
AWS	New Zealand	$4.5B	September 2021
AWS	Singapore	$9B	May 2024
AWS	Spain	$17B	July 2024
AWS	Taiwan	$5B	June 2025
AWS	Thailand	$5B	January 2025
Google	Belgium	$1.2B	April 2024
Google	Finland	$1.1B	May 2024
Google	Malaysia	$2B	May 2024
Google	Netherlands	$643M	April 2024
Google	Norway	$700M	May 2024
Google	United Kingdom	$1B	January 2024
Microsoft	Australia	$3.3B	October 2023
Microsoft	France	$4.3B	May 2024
Microsoft	Germany	$3.5B	February 2024
Microsoft	Indonesia	$1.7B	April 2024
Microsoft	Japan	$2.9B	April 2024
Microsoft	Kenya	$1B	May 2024
Microsoft	Malaysia	$2.2B	May 2024
Microsoft	Mexico	$1.3B	September 2024
Microsoft	Spain	$2.9B	June 2024
Microsoft	Sweden	$3.2B	June 2024
Microsoft	Switzerland	$400M	June 2025
Microsoft	United Kingdom	$3.2B	November 2023

of providing startups and researchers with access to the high-performance computing resources necessary for training AI models.

In Ferbuary 2025, the EU announced the InvestAI initiative, aiming to mobilize the much more significant sum of €200B for AI development (but still far short of the US's $500B Stargate project). The €50B in public funds and €150B from private sector commitments is to be used for five additional 'AI gigafactories' dedicated to development and training of next-generation AI models, each with over 100,000 processors, along with a number of other initiatives. The private sector partners are a diverse group of companies from both EU and, notably, non-EU countries.

https://doi.org/10.1017/nlp.2025.10007 Published online by Cambridge University Press

There has been a flood of interest in the InvestAI scheme: 76 expressions of interest have been submitted, covering 60 potential sites across 16 EU countries.

There's also activity at the member-state level. In addition to the earlier-mentioned multi-billion-dollar data centre builds by AWS, Google and Microsoft in various countries across Europe, there have been announcements of data centre builds in Finland, Germany, Italy, Spain and Sweden. Particularly notable is France's plan to build a €30-50B one-gigawatt data centre campus, backed by the UAE's MGX fund. In June of this year, Nvidia announced plans for 20 AI factories across Europe.

What's common to pretty much all of these endeavours is that at least part of the funding comes from outside the EU. If that means that a foreign investor controls key decisions, then sovereignty is undermined.

4.1.2 Chips

The EU also wants to establish a robust AI chip manufacturing ecosystem within Europe. The stated goal is to supply 20% of the world's semiconductor chips by 2030, a share reflecting the percentage of global end users that are located in Europe. But a report by the official European court of auditors found the EU's strategy to be 'deeply disconnected from reality'.

There are indeed fledgling attempts at domestic chip development: Spain's Openchip, founded in 2024, has plans to build a European AI chip that would rival those from Nvidia, and French startup SiPearl, which recently received €130m in Series A funding, similarly positions itself as a challenger to Nvidia. But it's hard to imagine either becoming anything more than a niche player.

Meanwhile, TSMC has announced plans to establish a design centre in Munich, as well as a manufacturing facility in Dresden through a joint venture called European Semiconductor Manufacturing Co (ESMC), and Intel has committed to a substantial expansion of its semiconductor operations across Europe, with an initial investment of at least €33B and the potential to increase that to €80B over the next decade.

So, once more, domestic footprints – which has to be an improvement – but overseas ownership.

4.1.3 Models

You can't say France and AI in the same sentence without also mentioning Mistral, Europe's best bet for a home-based competitor to OpenAI. The Paris-based AI lab, founded by Meta and Deepmind alumni, came out of stealth with €105 million in seed funding in mid 2023, aiming to develop AI technology that reflects European values and interests. It has since fielded multiple iterations of a range of AI models; at the time of writing it was in talks with MGX on a $1B equity financing round. The company was last valued at $6B in June 2024; that's tiny compared to OpenAI's current $300B valuation, but its models perform well on some key benchmarks, so it clearly punches above its weight class. The company has also announced plans to build its first data centre, and has launched a 24-billion-parameter language model optimized for Arabic content.

More broadly, OpenEuroLLM – a collaboration between 20 organizations – has launched a €37.4m EU-backed initiative to develop open-source language models covering all European languages by 2028. This includes the current 24 official EU languages, as well as languages for countries such as Albania that are currently negotiating for entry to the EU market.

Language models tuned to specific European languages are being developed in a number of countries in Europe, including Denmark, Italy, the Netherlands and Spain. Germany's Aleph Alpha, once a promising European contender in generative AI, has since pivoted to offering AI operational software in the face of escalating competition and internal challenges.

https://doi.org/10.1017/nlp.2025.10007 Published online by Cambridge University Press

4.2 The UK

At the beginning of 2025, the UK government unveiled a 50-point AI action plan, including dedicated 'growth zones' and £14B in private sector investments. Growth zones are designated areas of the UK that will have enhanced access to power and fast-tracking of planning approvals; local and regional authorities across the country were invited to bid to become one of these, with over 200 bids being received. Sites were required to have access, or plans for access, to power connections of at least 500MW; the government's goal is to encourage energy companies and data centre developers to provide the infrastructure needed.

Not long after Huang's earlier-mentioned criticism of the UK's lack of AI infrastructure, the UK pledged £ 2B in funding for its AI strategy, including £1B on building out AI compute by 2030 and another £750M for a new national supercomputer; this reversed cuts made to the previous government's promised AI funding less than a year earlier.

4.2.1 Data centres

We see a familiar pattern here, of overseas corporations investing in domestic data centre builds. A $13B data centre in North East England proposed by US private equity group Blackstone has been given the go-ahead by council planners; Oracle announced a $5B investment in UK cloud infrastructure and AI capabilities over five years; Japanese-owned Colt Data Services plans expansion that will make its Hayes campus one of the largest in the UK and Nvidia launched a major UK AI initiative with a £1.5B investment, including a new AI lab in Bristol, which is home to Isambard-AI, the UK's most powerful supercomputer with 5,000 Nvidia GPUs.

There are some UK data centre builds that appear to be funded entirely by UK companies: Apatura, a UK-based renewable energy developer, plans to invest £3.9B converting the former Ravenscraig steelworks into what would become the UK's largest green AI data centre; NScale plans to invest $2.5B in the UK's data centre industry; and Latos Date Centres aims to establish 40 AI-focused facilities by 2030.

4.2.2 Chips

In early 2023, the then-Conservative UK government announced an investment of up to £1B in the domestic semiconductor industry. Observers were quick to criticise the move as lacking ambition, drawing comparison with the US pledge of $52B (£42B) in subsidies for semiconductor manufacturing and research, and the EU's €43B (£37B) investment plan for the sector. The program still exists, but its future is unclear.

Later in the year, the government announced plans to spend £100M on chips from AMD, Intel and Nvidia in an effort to build a national AI resource. In the scheme of things, that's another small number.

4.2.3 Models

Google Deepmind, one of the leading frontier AI labs, was originally a UK company prior to being acquired by Google; it's still the UK's most prominent AI research lab.

Synthesia (valued at $2.1B), Stability AI (valued at $1B in mid 2024) and Poly AI (valued at close to $500M) are respected UK-based model developers, but again small in comparison to the US leaders. Much of Stability's activity has shifted to the US.

4.3 The Middle East

The Middle East's profile in the AI world is a little unusual. As well as having its own homegrown AI activity, with the UAE and Saudi Arabia battling it out to be the region's AI leader,

https://doi.org/10.1017/nlp.2025.10007 Published online by Cambridge University Press

Middle Eastern sovereign wealth funds – mainly from Saudi Arabia, UAE, Kuwait and Qatar – have been investing significant amounts in AI enterprises across the world as part of an economic diversification strategy. Abu Dhabi-based MGX, a dedicated AI fund, has been the most visible investor – it plans to invest $8B–$10B a year in AI, primarily in the US – but Saudi Arabia has also set aside $40B to invest in AI. MGX is a party to the AI Infrastructure Partnership, a global initiative launched in September 2024 to mobilize up to $100B in investments, with an initial target of $30B in private equity capital, to develop and expand AI data centers and the energy systems required to power them.

4.3.1 Data centres

AWS has partnered with Saudi Arabia's Humain to establish a $5B AI Zone, including dedicated AWS AI infrastructure, and the UAE's G42 is partnering with OpenAI, Oracle, Nvidia, SoftBank Group and Cisco to build Stargate UAE, a 1GW AI compute cluster within Abu Dhabi's new UAE-US AI Campus. Cerio, a Canadian firm, and Cloudiogram, a regional integrator, have partnered to deliver AI factory services across the Arab world.

4.3.2 Chips

Both the UAE and Saudi Arabia have been buying up chips to power their ambitions. By late 2024, concerns about the tech getting to China led to the Biden administration considering export limits, following its earlier banning of the sale of advanced AI chips to China. Similar concerns surfaced as a result of Trump's eagerness to see deals done in the first few months of his second term, but Trump's AI czar David Sacks defended that activity on the basis that China would step in if the US didn't engage.

Humain, an AI venture owned by Saudi Arabia's $925B Public Investment Fund, will buy several hundred thousand of Nvidia's AI chips over the next five years. Nvidia has reached a preliminary agreement with the UAE to supply 500,000 advanced AI chips annually from 2025, with 20% allocated to UAE's G42. Those chips will be used to build the largest AI campus outside the US.

TSMC and Samsung are discussing building Middle Eastern megafactories, with projects in the UAE potentially worth more than $100B.

4.3.3 Models

The Falcon-40B LLM from Abu Dhabi's Technology Innovation Institute (TII) knocked Meta's LLaMA off the top spot on the Open LLM Leaderboard when it was launched in mid-2023. The subsequently-released Falcon 180B was the largest available open-source language model until it was dethroned by Meta's LLaMa 3. TII has been a prolific provider of AI models since its release of Falcon 40B in early 2023; earlier this year, it launched Falcon Arabic, the first Arabic model in its Falcon series.

UAE's G42 has launched an open-source Arabic language model, developed in collaboration with Cerebras Systems, that contains 13 billion parameters and combines Arabic and English data. Microsoft is investing $1.5B into G42, whose AI apps and services will run on Microsoft Azure as part of the deal. This investment again raised concerns that it could transfer critical US tech abroad, but the White House supports the deal because it eliminates Huawei's influence.

4.4 Southeast Asia

4.4.1 Data centres

Microsoft has launched cloud regions in Indonesia and Malaysia, investing $3.9B to boost Southeast Asia's digital infrastructure, and Nvidia and Indosat are collaborating to build an AI center in Indonesia with a $200M investment.

https://doi.org/10.1017/nlp.2025.10007 Published online by Cambridge University Press

Google is investing $2B in Malaysia to establish its first data center and Cloud region. In May 2025, Malaysia announced a deal to power Malaysia's national AI system with Huawei's Ascend GPU servers; the planned deployment was quickly retracted following US pressure.

Singapore is investing over $743M in AI over five years to enhance capabilities and position itself as a global business and innovation hub. Salesforce has announced a $1B investment in Singapore, aiming to expand its AI capabilities.

California-based Fir Hills is partnering with South Korea to build a 3GW data centre, a $10B project expected to complete in 2028. Hewlett Packard Enterprise has been selected by the Korea Institute of Science and Technology Information to build South Korea's most powerful supercomputer, with 600 petaflops performance.

Nvidia and Foxconn have partnered with Taiwan's government to build an AI supercomputer featuring 10,000 Blackwell GPUs.

4.4.2 Chips

South Korea plans to invest $7B in AI, including a separate fund for AI semiconductor firms, to strengthen its position in cutting-edge semiconductor chips.

TSMC founder Morris Chang predicts a 'chip renaissance' in Japan as the country aims to restore its chipmaking excellence from the 1980s.

4.4.3 Models

GoTo Group and Indosat Ooredoo Hutchison have launched an enhanced version of Sahabat-AI, a 70-billion-parameter multilingual language model developed in Indonesia.

The Japanese government and major technology companies are investing in the development of a Japanese language model that aims to better represent cultural and linguistic subtleties compared to existing models like ChatGPT.

Singapore is leading a research initiative with a budget of around $53M to develop an LLM that caters to the diverse mix of culture and language in Southeast Asia. IBM and AI Singapore are collaborating to fine-tune SEA-LION, a Southeast Asian LLM developed by AISG to support 11 Southeast Asian languages, including Indonesian, Thai, Tamil, Filipino and Burmese.

Naver, South Korea's dominant internet search engine, will offer tailored versions of its latest ChatGPT-like model to foreign governments concerned about US data controls.

Taiwan is developing its own language model, Taide, to counter China's influence.

4.5 India

First, infrastructure: In March 2024, the Indian government launched the IndiaAI Mission with a budget of around $1.25B, focussed on developing indigenous AI technologies and establishing high-performance computing facilities to support AI workloads; a cornerstone of the effort is a planned supercomputer that will have at least 10,000 GPUs.

As a fellow member of BRICS, you might expect India to find a more natural ally in China, but here again the US hyperscalers are on the ground first. AWS plans to invest around $8.2B over the coming years; Microsoft is set to invest $3B over two years to expand its Azure cloud and AI infrastructure in India; and Google plans to set up an 8-story, 381,000 sq ft data center in Navi Mumbai. India is OpenAI's second-largest market, so it's not surprising that OpenAI has also begun talks to establish Indian data centres in partnership with Microsoft.

Although Chinese activity in India is limited – Alibaba Cloud ceased its Indian data center operations in 2024 as part of a strategic shift to focus on other markets – it's not entirely absent: Lenovo plans to manufacture AI servers in the country and has opened a new AI-centric R&D lab in Bengaluru.

https://doi.org/10.1017/nlp.2025.10007 Published online by Cambridge University Press

On the models front: In September 2024, the government launched BharatGen, a state-funded initiative to develop indigenous large language models. Several other LLM developments are underway: researchers at India's AI4Bharat raised $12M in seed funding to develop custom-made LLMs for India-centric use-cases; Tech Mahindra is developing an LLM called Project Indus that will be able to speak in 40 Indic languages; Krutrim launched India's first AI frontier research lab with a $230M investment; and Soket AI Labs is also developing frontier AI models in India, following DeepSeek's example of building advanced language models with limited resources.

4.6 Africa

Africa's AI sector's growth is hindered by inadequate funding, infrastructure challenges and regulatory conflicts. The US and China are competing to influence the development of the sector; China's implementation of a 2025–27 action plan has led to expectations of deeper Sino-African cooperation in technology, including digital infrastructure.

With regard to infrastructure, the usual suspects are, once more, already in place. In May 2024, Microsoft, in collaboration with UAE-based G42, announced a $1B investment to build a data centre in Kenya. Following earlier investments, Microsoft announced plans to invest around $300M in new data centre infrastructure in South Africa. AWS and Google also have a data centre presence in South Africa, and AWS has extended its plans to offer cloud computing in Morocco and Senegal, although so far these are not AI data centres.

There's some home-grown activity, but once again dependent on Western tech: Pan-African Cassava Technologies plans to build Africa's first AI factory, using Nvidia chips first at its data centres in South Africa, and subsequently at its other facilities in Egypt, Kenya, Morocco and Nigeria; and Africa Data Centres announced in 2021 a $500M plan to build 10 interconnected hyperscale data centers across Africa, although this project appears to be considerably behind schedule.

With regards to models: Lelapa AI, a South African startup, has developed InkubaLM, a multilingual LLM designed for five widely spoken African languages: Swahili, Yoruba, isiXhosa, Hausa and isiZulu. Kenya-based Jacaranda Health has launched UlizaLlama, an open-access LLM tailored for Swahili. EqualyzAI, a startup based in Nigeria, has launched Africa-centric AI models that understand local languages and cultural contexts.

5. Summing up

So, given all that, how are aspirations to AI sovereignty faring?

What we earlier called full-stack sovereignty – independence at every level of the AI technical stack – is unachievable for all except the US and China, principally because of domination of the AI chip market. Instead, every other nation in the world has to be satisfied with some form of 'partial sovereignty'. In the best case, you'll have domestically-funded data centres that afford you complete control; more likely you'll have data centres with a mix of ownership structures and their corresponding risks; and if you're a smaller, poorer nation, perhaps the best you can hope for are foreign-operated data centres populated by foreign-made chips but hosting locally-made or fine-tuned AI models.

With regard to chips, you might take the view that it doesn't matter who makes them: once you've bought them, they are yours to do with as you wish. But while there appears to be no evidence that Nvidia chips have remote kill switches, that doesn't mean they are safe from interference. It's feasible that your chips could be restricted via driver updates or software locks. And in any case, they'll be outdated in a couple of years. When replacement comes due, you'll already be locked into the manufacturer's firmware and software infrastructure, and, as China does today, you'll just have to take the best that export controls at the time permit.

https://doi.org/10.1017/nlp.2025.10007 Published online by Cambridge University Press

What about data centres? You might argue that, regardless of ownership, at least your nation's data is safe. After all, that's the central selling point the hyperscalers use to promote their wares: having built their data centres on your territory, you keep that data within your national borders, free from misuse and appropriation. Well, maybe. Just bear in mind that the US Clarifying Lawful Overseas Use of Data Act (Cloud Act), enacted in 2018, provides a legal framework for US law enforcement agencies to access electronic data held by US-based technology companies, regardless of where the data is physically stored.

So, your scope for true sovereignty may ultimately be limited to the domestic development of culturally and linguistically appropriate AI models, assuming you can find enough digital data to train or fine-tune the models and have the budget for the training runs required – resources that might well be in short supply in smaller countries. That aspect of sovereignty is certainly not to be sniffed at; you could make an argument that it is the most important aspect in terms of a country's cultural uniqueness and independence. But, ultimately, you possess that sovereignty, and are only able to exercise it, at the whim of external forces.

Meanwhile, every GPU sold is an immediate increment to Nvidia's bottom line, and it's certainly on a roll: the company recently became the first publicly traded company to achieve a $4 trillion market valuation.

The hyperscalers, on the other hand, are making big investments with the expectation of returns over the longer term. Those expectations may not be realised if it turns out we're just in an AI bubble, and that bubble bursts.

But either way, in the world of sovereign AI, Jensen Huang looks to be king.

If you'd like to keep up to date with everything that's happening in the NLP industry, consider subscribing to the comprehensive and free *This Week in NLP* newsletter at https://thisweekinnlp.substack.com/.

Cite this article: Dale R (2025). Sovereign AI in 2025. *Natural Language Processing* **31**, 1312–1321. https://doi.org/10.1017/nlp.2025.10007

https://doi.org/10.1017/nlp.2025.10007 Published online by Cambridge University Press